Vol. 41, Parts 1 and 2 June 1954

BIOMETRIKA

FOUNDED BY
W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

MANAGING EDITOR
E. S. PEARSON

ASSOCIATE EDITORS
M. G. KENDALL JOHN WISHART

in consultation with
HARALD CRAMÉR R. C. GEARY
J. B. S. HALDANE D. G. KENDALL

ISSUED BY
THE BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON
PRINTED AT THE UNIVERSITY PRESS, CAMBRIDGE

[*Issued 3 June 1954*]

VOLUME 41, PARTS 1 AND 2 JUNE 1954

DISTRIBUTION THEORY OF TWO ESTIMATES FOR STANDARD DEVIATION BASED ON SECOND VARIATE DIFFERENCES*

BY A. R. KAMAT
University College, London

1. INTRODUCTION

1·1. *Preliminary remarks*

Anderson (1927) proposed the use of variate differences of various orders to eliminate the effects of a polynomial trend. Tintner (1940) has given a systematic account of the same in his book. Von Neumann, Kent, Bellinson & Hart (1941) advocated the use of the mean-square successive difference $\delta^2 = \sum_{i=1}^{n-1} (x_i - x_{i+1})^2/(n-1)$, that is, an estimate based on the first variate difference, to estimate σ^2 when the mean of the parent population is undergoing a slow-moving continuous trend. Recently the present author (1953*b*) has discussed the approximate distribution of another estimate based on the first variate difference, viz. the mean successive difference $d = \sum_{i=1}^{n-1} |x_i - x_{i+1}|/(n-1)$ which is useful under the same circumstances. In this paper we deal with the approximate distribution of the following two estimates based on the second variate difference:

(1) the mean square successive second difference:

$$\delta_2^2 = \frac{1}{n-2} \sum_{i=1}^{n-2} (x_i - 2x_{i+1} + x_{i+2})^2 = \frac{1}{n-2} \Sigma (\Delta^2 x_i)^2;$$

(2) the mean successive second difference:

$$d_2 = \frac{1}{n-2} \sum_{i=1}^{n-2} |x_i - 2x_{i+1} + x_{i+2}| = \frac{1}{n-2} \Sigma |\Delta^2 x_i|.$$

If the variables x_i are distributed normally and independently with a common mean and variance, σ^2, then the variate difference estimators of the variance are distinctly less efficient than the usual estimator $s^2 = \sum_i (x_i - \bar{x})^2/(n-1)$. In fact, it has been shown (Morse & Grubbs, 1947; Guest, 1951; Kamat, 1953*c*) that the asymptotic efficiencies are

$$\delta^2: \ 66{\cdot}7\,\%, \qquad \delta_2^2: \ 51{\cdot}4\,\%, \qquad\qquad d: \ 60{\cdot}5\,\%, \qquad d_2: \ 47{\cdot}1\,\%.$$

Variate difference methods may be valuable, however, in cases where it is reasonable to suppose that independent normal deviations with distribution $N(0, \sigma)$ are superposed upon a trend, and we wish either to estimate the value of σ or, perhaps, to test whether it has changed from one series of observations to another. In such cases we suppose that $x_i = \mu_i + z_i$, where z_i is $N(0, \sigma)$ and μ_i is a slow-moving trend function. If the trend function, μ_i, may be 'locally' represented by a straight line or a parabola with moderate curvature the first- or second-order variate difference estimators will be free of the heavy bias that affects the estimator s^2.

Different aspects of this subject have been discussed by a number of authors (Morse & Grubbs, 1947; Keen & Page, 1953). As an illustration of the extent to which the second variate differences succeed in eliminating trend, we may quote three examples given by

* This paper is part of a thesis approved for the Ph.D. degree of the University of London.

Tintner. The figures quoted below are values of the variances estimated from the squares of variate differences of various orders for: (1) annual wheat-flour prices (Tintner, 1952, p. 313); (2) annual wool prices (Tintner, 1940, p. 70); (3) annual raw-silk prices (Tintner, 1940, p. 71). All data are for U.S.A. for the 48 years 1890–1937:

p = order of variate difference	Ex. (1)	Ex. (2)	Ex. (3)
0	4·7969	0·1069	2·8914
1	0·7020	0·0277	0·3849
2	0·4402	0·0259	0·2714
3	0 3931	0·0262	0·2317
4	0·3767	0·0264	0·2033
5	0·3662	0·0263	0·1824

Tintner has shown that the estimated variance may be regarded as stabilized at $p = 1$ or 2 for Ex. (1), at $p = 1$ or 2 for Ex. (2) and at $p = 2$ for Ex. (3).

As stated above, the object of the present paper is primarily to derive the higher moments of the sampling distributions of δ_2^2 and d_2 and indicate their approximate distributions. In a further paper it is hoped to discuss various practical uses of distributions based on these moments and other allied problems.

1·2. *Notation*

Let x_i $(i = 1, 2, \ldots, n)$ denote a sequence of observations from a normal population with a constant mean and S.D. σ. Then the second variate difference is defined as

$$\Delta^2 x_i = x_i - 2x_{i+1} + x_{i+2},$$

and the two estimates discussed below are:

$$\delta_2^2 = \sum_{i=1}^{n-2} (\Delta^2 x_i)^2/(n-2) \quad \text{and} \quad d_2 = \sum_{i=1}^{n-2} |\Delta^2 x_i|/(n-2). \tag{1}$$

For simplification we shall use the following notation. Let $\sigma' = \sigma(\Delta^2 x_i) = \sqrt{6}\,\sigma$, then we define

$$z_i = \Delta^2 x_i/(\sqrt{6}\,\sigma) = \Delta^2 x_i/\sigma' \quad (i = 1, 2, \ldots, n-2), \tag{2}$$

so that

$$\left.\begin{aligned} \sigma(z_i) &= 1 \\ \text{and} \quad \rho_1 = \rho(z_i, z_{i+1}) &= \rho(\Delta^2 x_i, \Delta^2 x_{i+1}) = -\tfrac{2}{3}, \\ \rho_2 = \rho(z_i, z_{i+2}) &= \rho(\Delta^2 x_i, \Delta^2 x_{i+2}) = \tfrac{1}{6}, \\ \rho_m = \rho(z_i, z_{i+m}) &= \rho(\Delta^2 x_i, \Delta^2 x_{i+m}) = 0 \quad (m > 2). \end{aligned}\right\} \tag{3}$$

2. The distribution theory of the mean-square successive second difference, δ_2^2

2·1. *Preliminary results*

Using the z variables defined in (2) above we have

$$\left.\begin{aligned} \mathscr{E}(\Sigma(\Delta^2 x_i)^2) &= \{\mathscr{E}(\Sigma z_i^2)\}\,\sigma'^2, \\ \mathscr{E}\{\Sigma(\Delta^2 x_i)^2\}^2 &= \{\mathscr{E}(\Sigma z_i^2)^2\}\,\sigma'^4, \quad \text{etc.} \end{aligned}\right\} \tag{4}$$

In order to obtain the first four moments of δ_2^2 we have therefore to evaluate the following expectations which can be found from the moments of multivariate normal distributions up to the four-variate case by substituting the appropriate values of ρ_m ($m = 1, 2$ and > 2):

$$
\left.\begin{array}{lll}
\mathscr{E}(z_1^2) = 1, & & \\
\mathscr{E}(z_1^4) = 3, & \mathscr{E}(z_1^2 z_2^2) = \frac{17}{9}, & \mathscr{E}(z_1^2 z_3^2) = \frac{19}{18}, \\
\mathscr{E}(z_1^6) = 15, & \mathscr{E}(z_1^4 z_2^2) = \frac{25}{3}, & \mathscr{E}(z_1^4 z_3^2) = \frac{10}{3}, \\
\mathscr{E}(z_1^2 z_2^2 z_3^2) = \frac{185}{54}, & \mathscr{E}(z_1^2 z_2^2 z_4^2) = \frac{35}{18}, & \mathscr{E}(z_1^2 z_3^2 z_5^2) = \frac{10}{9}, \\
\mathscr{E}(z_1^8) = 105, & \mathscr{E}(z_1^6 z_2^2) = 55, & \mathscr{E}(z_1^6 z_3^2) = \frac{35}{2}, \\
\mathscr{E}(z_1^4 z_2^4) = \frac{1235}{27}, & \mathscr{E}(z_1^4 z_3^4) = \frac{595}{54}, & \mathscr{E}(z_1^4 z_2^2 z_3^2) = \frac{410}{27}, \\
\mathscr{E}(z_1^2 z_2^4 z_3^2) = \frac{1195}{54}, & \mathscr{E}(z_1^4 z_2^2 z_4^2) = \frac{17}{2}, & \mathscr{E}(z_1^2 z_2^4 z_4^2) = \frac{242}{27}, \\
\mathscr{E}(z_1^2 z_2^2 z_4^4) = 6, & \mathscr{E}(z_1^4 z_3^2 z_5^2) = \frac{7}{2}, & \mathscr{E}(z_1^2 z_3^4 z_5^2) = \frac{199}{54}, \\
\mathscr{E}(z_1^2 z_2^2 z_3^2 z_4^2) = \frac{643}{108}, & \mathscr{E}(z_1^2 z_3^2 z_4^2 z_5^2) = \frac{587}{162}, & \\
\mathscr{E}(z_1^2 z_3^2 z_4^2 z_5^2) = \frac{286}{81}, & \mathscr{E}(z_1^2 z_3^2 z_4^2 z_6^2) = \frac{649}{324}, & \\
\mathscr{E}(z_1^2 z_3^2 z_5^2 z_6^2) = \frac{166}{81}, & \mathscr{E}(z_1^2 z_3^2 z_5^2 z_7^2) = \frac{379}{324}. &
\end{array}\right\} \quad (5)
$$

2·2. *Moments of* δ_2^2

Let $n-2 = m$ and $\sum_{i=1}^{n-2} z_i^2 = Z$, then

$$
\left.\begin{aligned}
\mathscr{E}(Z) &= m\mathscr{E}(z_1^2), \\
\mathscr{E}(Z^2) &= m\mathscr{E}(z_1^4) + 2(m-1)\,\mathscr{E}(z_1^2 z_2^2) + 2(m-2)\,\mathscr{E}(z_1^2 z_3^2) \\
&\quad + (m-2)(m-3)\,\mathscr{E}^2(z_1^2), \\
\mathscr{E}(Z^3) &= m\mathscr{E}(z_1^6) + 6(m-1)\,\mathscr{E}(z_1^4 z_2^2) + 6(m-2)\,\mathscr{E}(z_1^4 z_3^2) \\
&\quad + 3(m-2)(m-3)\,\mathscr{E}(z_1^4)\,\mathscr{E}(z_1^2) + 6(m-2)\,\mathscr{E}(z_1^2 z_2^2 z_3^2) \\
&\quad + 12(m-3)\,\mathscr{E}(z_1^2 z_2^2 z_4^2) + 6(m-4)\,\mathscr{E}(z_1^2 z_3^2 z_5^2) \\
&\quad + 6(m-3)(m-4)\,\mathscr{E}(z_1^2 z_2^2)\,\mathscr{E}(z_1^2) + 6(m-4)(m-5)\,\mathscr{E}(z_1^2 z_3^2)\,\mathscr{E}(z_1^2) \\
&\quad + (m-4)(m-5)(m-6)\,\mathscr{E}^3(z_1^2), \\
\mathscr{E}(Z^4) &= m\mathscr{E}(z_1^8) + 8(m-1)\,\mathscr{E}(z_1^6 z_2^2) + 8(m-2)\,\mathscr{E}(z_1^6 z_3^2) \\
&\quad + 4(m-2)(m-3)\,\mathscr{E}(z_1^6)\,\mathscr{E}(z_1^2) + 6(m-1)\,\mathscr{E}(z_1^4 z_2^4) \\
&\quad + 6(m-2)\,\mathscr{E}(z_1^4 z_3^4) + 3(m-2)(m-3)\,\mathscr{E}^2(z_1^4) \\
&\quad + 24(m-2)\,\mathscr{E}(z_1^4 z_2^2 z_3^2) + 12(m-2)\,\mathscr{E}(z_1^2 z_2^4 z_3^2) \\
&\quad + 24(m-3)\,\mathscr{E}(z_1^4 z_2^2 z_4^2) + 24(m-3)\,\mathscr{E}(z_1^2 z_2^4 z_4^2) \\
&\quad + 24(m-3)\,\mathscr{E}(z_1^2 z_2^2 z_4^4) + 24(m-4)\,\mathscr{E}(z_1^4 z_3^2 z_5^2) \\
&\quad + 12(m-4)\,\mathscr{E}(z_1^2 z_3^4 z_5^2) + 24(m-3)(m-4)\,\mathscr{E}(z_1^4 z_2^2)\,\mathscr{E}(z_1^2) \\
&\quad + 24(m-4)(m-5)\,\mathscr{E}(z_1^4 z_3^2)\,\mathscr{E}(z_1^2) + 12(m-3)(m-4)\,\mathscr{E}(z_1^2 z_2^2)\,\mathscr{E}(z_1^4) \\
&\quad + 12(m-4)(m-5)\,\mathscr{E}(z_1^2 z_3^2)\,\mathscr{E}(z_1^4) + 6(m-4)(m-5)(m-6)\,\mathscr{E}(z_1^4)\,\mathscr{E}^2(z_1^2) \\
&\quad + 24(m-3)\,\mathscr{E}(z_1^2 z_2^2 z_3^2 z_4^2) + 24(m-4)\,\mathscr{E}(z_1^2 z_2^2 z_4^2 z_5^2) \\
&\quad + 48(m-4)\,\mathscr{E}(z_1^2 z_3^2 z_4^2 z_5^2) + 24(m-5)\,\mathscr{E}(z_1^2 z_3^2 z_4^2 z_6^2) \\
&\quad + 48(m-5)\,\mathscr{E}(z_1^2 z_2^2 z_4^2 z_6^2) + 24(m-6)\,\mathscr{E}(z_1^2 z_3^2 z_5^2 z_7^2) \\
&\quad + 24(m-4)(m-5)\,\mathscr{E}(z_1^2 z_2^2 z_3^2)\,\mathscr{E}(z_1^2) \\
&\quad + 48(m-5)(m-6)\,\mathscr{E}(z_1^2 z_2^2 z_4^2)\,\mathscr{E}(z_1^2) \\
&\quad + 24(m-6)(m-7)\,\mathscr{E}(z_1^2 z_3^2 z_5^2)\,\mathscr{E}(z_1^2) \\
&\quad + 12(m-4)(m-5)\,\mathscr{E}^2(z_1^2 z_2^2) + 24(m-5)(m-6)\,\mathscr{E}(z_1^2 z_2^2)\,\mathscr{E}(z_1^2 z_3^2) \\
&\quad + 12(m-6)(m-7)\,\mathscr{E}^2(z_1^2 z_3^2) \\
&\quad + 12(m-5)(m-6)(m-7)\,\mathscr{E}(z_1^2 z_2^2)\,\mathscr{E}^2(z_1^2) \\
&\quad + 12(m-6)(m-7)(m-8)\,\mathscr{E}(z_1^2 z_3^2)\,\mathscr{E}^2(z_1^2) \\
&\quad + (m-6)(m-7)(m-8)(m-9)\,\mathscr{E}^4(z_1^2).
\end{aligned}\right\} \quad (6)
$$

Substituting the values of various terms in (6) from (5), after considerable simplifications, we have

$$\left.\begin{aligned}\mu_1'(Z) &= m,\\ \mu_2'(Z) &= m^2+\tfrac{35}{9}m-2,\\ \mu_3'(Z) &= m^3+\tfrac{35}{3}m^2+\tfrac{254}{9}m-\tfrac{280}{9},\\ \mu_4'(Z) &= m^4+\tfrac{70}{3}m^3+\tfrac{4597}{27}m^2+\tfrac{2750}{9}m-\tfrac{15664}{27}.\end{aligned}\right\} \tag{7}$$

For $\delta_2^2 = Z\sigma'^2/m$, therefore, we obtain the moments

$$\left.\begin{aligned}\mu_1' &= \sigma'^2, & \mu_3 &= \frac{308m-280}{9m^3}\sigma'^6,\\ \mu_2 &= \frac{35m-18}{9m^2}\sigma'^4, & \mu_4 &= \frac{1}{27m^4}(1225m^2+11610m-15664)\sigma'^8.\end{aligned}\right\} \tag{8}$$

Since $\sigma'^2 = 6\sigma^2$, still writing $m = n-2$, we have for the statistic δ_2^2/σ^2,

$$\left.\begin{aligned}\mu_1' &= 6, & \beta_1 &= 9(308m-280)^2/(35m-18)^3,\\ \mu_2 &= 4(35m-18)/m^2, & \beta_2 &= 3+3(12870m-15988)/(35m-18)^2.\end{aligned}\right\} \tag{9}$$

It is to be noted that the formulae for μ_3 and β_1 hold for $m \geqslant 4$ ($n \geqslant 6$), and those for μ_4 and β_2 hold for $m \geqslant 6$ ($n \geqslant 8$). The characteristic equation of the matrix of the quadratic form $\Sigma(\Delta^2 x_i)^2$, the derivation of which is considered in the last section, provides a check on the long algebra of the evaluation of the moments.

2·3. *The distribution of* δ_2^2

Except in the very exceptional case of $n = 4$ which is discussed below in § 2·4 it seems difficult to obtain the exact distribution of δ_2^2. Table 1 gives moment constants for δ_2^2 for $n = 5, 7, 10, 15, 20, 25, 30, 40, 50$. The approach to normality is slow and the β_1, β_2 points lie in the Pearson Type VI region, which suggests that a curve of this type may provide a good approximation to the distribution.

Table 1. *Standard deviation and* β_1, β_2 *values for* δ_2^2/σ^2

n	$\sigma(\delta_2^2/\sigma^2)$	β_1	β_2
5	6·2183	5·6684	11·9679
7	5·0120	3·6922	8·8861
10	4·0446	2·3870	6·8010
15	3·2161	1·4956	5·3772
20	2·7487	1·0880	4·7275
25	2·4394	0·8548	4·3563
30	2·2154	0·7038	4·1163
40	1·9064	0·5201	3·8245
50	1·6987	0·4124	3·6536

Note. The mean of δ_2^2/σ^2 is 6.

2·4. *Exact distribution of δ_2^2 for $n = 4$*

For $n = 4$ the exact distribution of δ_2^2 can be obtained by a method similar to the one by which von Neumann *et al.* (1941) have obtained the exact distribution of the mean-square successive difference δ^2 for $n = 3$. In this case

$$\Pr\left\{\frac{\delta_2^2}{\sigma^2} \leqslant \delta_0^2\right\} = \Pr\left\{\frac{1}{2\sigma^2}((\Delta^2 x_1)^2 + (\Delta^2 x_2)^2) \leqslant \delta_0^2\right\}$$

$$= \Pr\{z_1^2 + z_2^2 \leqslant \tfrac{1}{3}\delta_0^2\}, \tag{10}$$

where $z_i = \Delta^2 x_i/(\sqrt{6}\sigma) \quad (i = 1, 2); \quad \sigma(z_i) = 1, \quad \rho(z_1 z_2) = -\tfrac{2}{3}.$

Now $$\Pr\{z_1^2 + z_2^2 \leqslant \tfrac{1}{3}\delta_0^2\} = \frac{3}{2\pi\sqrt{5}} \iint\limits_{z_1^2 + z_2^2 < \frac{1}{3}\delta_0^2} \exp\{-\tfrac{9}{10}(z_1^2 + z_2^2 + \tfrac{4}{3}z_1 z_2)\}\, dz_1 dz_2.$$

Normalizing the quadratic form, this integral can be written as

$$\frac{3}{2\pi\sqrt{5}} \iint\limits_{z_1'^2 + z_2'^2 < \frac{1}{3}\delta_0^2} \exp\{-\tfrac{3}{2}(z_1'^2 + \tfrac{1}{5}z_2'^2)\}\, dz_1' dz_2' = \frac{3}{2\pi\sqrt{5}} \int_0^{\delta_0/\sqrt{3}} r\, e^{-\frac{1}{2}r^2} \left(\int_0^{2\pi} e^{\frac{1}{5}r^2 \sin^2\theta}\, d\theta\right) dr. \tag{11}$$

The substitution $\frac{6}{5}r^2 = -2iu$, $\phi = \frac{1}{2}\pi - 2\theta$ transforms the integral

$$\int_0^{2\pi} e^{\frac{1}{5}r^2 \sin^2\theta}\, d\theta = 2\pi\, e^{-iu} J_0(u),$$

where $J_0(u)$ is the Bessel function of the zero order. Therefore, finally,

$$\Pr\left\{\frac{\delta_2^2}{\sigma^2} \leqslant \delta_0^2\right\} = \frac{3}{\sqrt{5}} \int_0^{\delta_0/\sqrt{3}} r\, e^{-\frac{3}{2}r^2} J_0(\tfrac{3}{5}ir^2)\, dr, \tag{12}$$

and the probability density function is given by

$$p(\delta_2^2/\sigma^2) = \frac{1}{2\sqrt{5}}\, e^{-\frac{3}{10}\frac{\delta_2^2}{\sigma^2}} J_0\left(\frac{i}{5}\frac{\delta_2^2}{\sigma^2}\right). \tag{13}$$

3. THE DISTRIBUTION THEORY OF THE MEAN SUCCESSIVE SECOND DIFFERENCE d_2

3·1. *Preliminary results*

Using the z-variables defined in § 1·2 above we have

$$\left.\begin{aligned} \mathscr{E}(\Sigma\,|\,\Delta^2 x_i\,|) &= \{\mathscr{E}(\Sigma\,|\,z_i\,|)\}\,\sigma', \\ \mathscr{E}(\Sigma\,|\,\Delta^2 x_i\,|)^2 &= \{\mathscr{E}(\Sigma\,|\,z_i\,|)^2\}\,\sigma'^2, \quad \text{etc.} \end{aligned}\right\} \tag{14}$$

In order to evaluate the first four moments of d_2 we require the values of the following expectations in addition to some of the formulae given in (5) above. They are found by substituting the appropriate values of ρ_m in the absolute moments given in Kamat (1953*a*). The last six expectations, viz. those of the form $\mathscr{E}(|\,z_i\,|\,|\,z_j\,|\,|\,z_k\,|\,|\,z_l\,|)$, have been evaluated from the expansion for the absolute moment (1, 1, 1, 1) given in the same paper. The values of $\mathscr{E}(|\,z_i\,|\,|\,z_j\,|\,|\,z_k\,|\,|\,z_l\,|)$ given below are believed to be correct to the last figure in each case.

$$\mathscr{E}(|z_1|) = \sqrt{\frac{2}{\pi}}, \quad \mathscr{E}(|z_1 z_2|) = \frac{2}{\pi}\left(\frac{\sqrt{5}}{3} + \frac{2\theta_1}{3}\right), \quad \mathscr{E}(|z_1 z_3|) = \frac{2}{\pi}\tfrac{1}{6}(\sqrt{(35)} + \theta_2),$$

$$\mathscr{E}(|z_1^3|) = 2\sqrt{\frac{2}{\pi}}, \quad \mathscr{E}(|z_1^2 z_2|) = \sqrt{\frac{2}{\pi}}\frac{13}{9}, \quad \mathscr{E}(|z_1^2 z_3|) = \sqrt{\frac{2}{\pi}}\frac{37}{36},$$

$$\mathscr{E}(|z_1 z_2 z_3|) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}}\left(\frac{5\sqrt{3}}{18} - \frac{11\pi}{108} + \frac{14\phi_1}{9}\right),$$

$$\mathscr{E}(|z_1 z_2 z_4|) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}}\left(\frac{\sqrt{(19)}}{6} + \tfrac{1}{6}\phi_2 + \tfrac{8}{3}\phi_3 - \tfrac{1}{6}\phi_4\right),$$

$$\mathscr{E}(|z_1 z_3 z_5|) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}}\left(\frac{\sqrt{(34)}}{6} + \tfrac{1}{3}\phi_5 - \tfrac{1}{36}\phi_6\right),$$

$$\mathscr{E}(|z_1^3 z_2|) = \frac{2}{\pi}\left(\frac{22\sqrt{5}}{27} + 2\theta_1\right), \quad \mathscr{E}(|z_1^3 z_3|) = \frac{2}{\pi}\left(\frac{73\sqrt{(35)}}{216} + \tfrac{1}{2}\theta_2\right),$$

$$\mathscr{E}(|z_1^2 z_2 z_3|) = \frac{2}{\pi}\left(\frac{53\sqrt{5}}{108} + \tfrac{8}{3}\theta_1\right), \quad \mathscr{E}(|z_1 z_2^2 z_3|) = \frac{2}{\pi}\left(\frac{17\sqrt{(35)}}{54} + \frac{19\theta_2}{18}\right),$$

$$\mathscr{E}(|z_1^2 z_2 z_4|) = \frac{2}{\pi}\left(\frac{13\sqrt{(35)}}{54} + \tfrac{1}{6}\theta_2\right), \quad \mathscr{E}(|z_1 z_2^2 z_4|) = \frac{2}{\pi}\frac{53}{36},$$

$$\mathscr{E}(|z_1 z_2 z_4^2|) = \frac{2}{\pi}\left(\frac{37\sqrt{5}}{108} + \tfrac{8}{3}\theta_1\right), \quad \mathscr{E}(|z_1^2 z_3 z_5|) = \frac{2}{\pi}\left(\frac{37\sqrt{(35)}}{216} + \tfrac{1}{6}\theta_2\right),$$

$$\mathscr{E}(|z_1 z_3^2 z_5|) = \frac{2}{\pi}\frac{19}{18},$$

where
$$\theta_1 = \sin^{-1}\tfrac{2}{3}, \qquad \theta_2 = \sin^{-1}\tfrac{1}{6},$$

$$\phi_1 = \sin^{-1}\frac{2}{\sqrt{7}}, \quad \phi_2 = \sin^{-1}\frac{1}{2\sqrt{5}}, \quad \phi_3 = \sin^{-1}\frac{4}{\sqrt{(35)}},$$

$$\phi_4 = \sin^{-1}\frac{2}{5\sqrt{7}}, \quad \phi_5 = \sin^{-1}\frac{1}{\sqrt{(35)}}, \quad \phi_6 = \sin^{-1}\tfrac{1}{35},$$

and
$$\mathscr{E}(|z_1 z_2 z_3 z_4|) = \left(\frac{2}{\pi}\right)^2 1{\cdot}80,$$

$$\mathscr{E}(|z_1 z_2 z_3 z_5|) = \left(\frac{2}{\pi}\right)^2 1{\cdot}51,$$

$$\mathscr{E}(|z_1 z_2 z_4 z_5|) = \left(\frac{2}{\pi}\right)^2 1{\cdot}524,$$

$$\mathscr{E}(|z_1 z_3 z_4 z_6|) = \left(\frac{2}{\pi}\right)^2 1{\cdot}253,$$

$$\mathscr{E}(|z_1 z_2 z_4 z_6|) = \left(\frac{2}{\pi}\right)^2 1{\cdot}2590,$$

$$\mathscr{E}(|z_1 z_3 z_5 z_7|) = \left(\frac{2}{\pi}\right)^2 1{\cdot}04157. \qquad (15)$$

3·2. *Moments of* d_2

Again writing $m = n-2$ and putting $Y = \sum_{i=1}^{n-2} |z_i|$, we can expand $\mathscr{E}(Y)$, $\mathscr{E}(Y^2)$, etc., into expressions which are obtained from the right-hand side of (6) by replacing z_i^{2k} by $|z_i^k|$ $(i = 1, 2, 3, 4, 5, 6, 7;\ k = 1, 2, 3, 4)$. Substituting in the resulting expressions the values of the various terms from (5) and (15), after considerable simplification, we finally get the first four moments of d_2 as follows:

$$\mu_1' = \sqrt{\frac{2}{\pi}}\sigma' = 1{\cdot}954410\sigma, \tag{16}$$

$$\mu_2 = (1{\cdot}062321m^{-1} - 0{\cdot}519368m^{-2})\frac{2}{\pi}\sigma'^2$$

$$= (4{\cdot}057769m^{-1} - 1{\cdot}983839m^{-2})\,\sigma^2, \tag{17}$$

$$\mu_3 = (1{\cdot}988102m^{-1} - 1{\cdot}666781m^{-2})\left(\frac{2}{\pi}\right)^{\frac{3}{2}}\sigma'^3, \tag{18}$$

$$\mu_4 = \{3(1{\cdot}062321)^2 m^{-2} + 1{\cdot}494065m^{-3} - 4{\cdot}864367m^{-4}\}\left(\frac{2}{\pi}\right)^2\sigma'^4. \tag{19}$$

From (17), (18) and (19) we have

$$\left.\begin{aligned}\beta_1 &= \frac{(1{\cdot}988102m - 1{\cdot}666781)^2}{(1{\cdot}062321m - 0{\cdot}519368)^3}\\ \text{and}\quad \beta_2 &= 3 + \frac{4{\cdot}8045m - 5{\cdot}6736}{(1{\cdot}062321m - 0{\cdot}519368)^2}.\end{aligned}\right\} \tag{20}$$

It may be noted that there is hardly any point in retaining more than three or four places of decimals in the coefficients in β_2 because of the limited accuracy in the values of $\mathscr{E}(|z_1 z_2 z_3 z_4|)$, etc., given at the end of § 3·1. The expressions for μ_3 and β_1 hold for $m \geqslant 4$ $(n \geqslant 6)$ and those for μ_4 and β_2 for $m \geqslant 6$ $(n \geqslant 8)$.

3·3. *The distribution of* d_2

Except in the very exceptional case of $n = 4$, where the exact distribution can be obtained as shown below in § 3·4, it seems difficult to obtain the exact distribution of d_2. Table 2

Table 2. *Standard deviation and* β_1, β_2 *values for* d_2

n	$\sigma(d_2/\sigma)$	β_1	β_2
5	1·0640	0·972	4·225
7	0·8557	0·622	3·799
10	0·6901	0·399	3·515
15	0·5481	0·249	3·322
20	0·4683	0·181	3·234
25	0·4155	0·142	3·183
30	0·3773	0·117	3·151
40	0·3247	0·086	3·111
50	0·2893	0·068	3·088

The mean of d_2/σ is 1·954410 for all values of n. The last figure of β_2 for $n \geqslant 10$ is not reliable.

gives moment constants for d_2 for $n = 5, 7, 10, 15, 20, 25, 30, 40, 50$. The (β_1, β_2) points tend to the normal point far more quickly than those for δ_2^2 (see Table 1) and lie in the Pearson Type I region.

3·4. *Exact distribution of d_2 for $n = 4$*

In this case following the method given by Kamat (1953*b*) for d for $n = 3$, we have

$$\Pr\{d_2/\sigma < d_0\} = \Pr\left\{\frac{1}{2\sigma}(|\Delta^2 x_1| + |\Delta^2 x_2|) \leqslant d_0\right\}$$

$$= \Pr\{|z_1| + |z_2| \leqslant \sqrt{\tfrac{2}{3}}\, d_0\}, \tag{21}$$

where $z_i = \Delta^2 x_i/(\sqrt{6}\sigma)$ $(i = 1, 2)$; and $\sigma(z_i) = 1$, $\rho(z_1, z_2) = -\frac{2}{3}$. Therefore

$$\Pr\{d_2/\sigma \leqslant d_0\} = 2\left\{\iint\limits_{x+y \leqslant \sqrt{\frac{3}{2}}\, d_0} f(x, y; \tfrac{2}{3})\, dx\, dy + \iint\limits_{x+y \leqslant \sqrt{\frac{3}{2}}\, d_0} f(x, y; -\tfrac{2}{3})\, dx\, dy\right\}, \quad (x, y > 0) \tag{22}$$

where $$f(x, y; \rho) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 + y^2 - 2\rho xy)\right].$$

The expression on the right-hand side of (22) can be transformed by linear transformations into $p_\rho(h, k)$ functions defined by

$$p_\rho(h, k) = \int_h^\infty \int_k^\infty f(x, y; \rho)\, dx\, dy,$$

which are given in Tables VIII and IX of Karl Pearson's (1931) *Tables for Statisticians and Biometricians*, Part II. The probability integral (21) can therefore be expressed as

$$\Pr\{d_2/\sigma \leqslant d_0\} = 1 - 2p_{\sqrt{\frac{1}{2}}}(d_0, 0) - 2p_{\sqrt{\frac{1}{2}}}(d_0/\sqrt{5}, 0) + 2p_{-\sqrt{\frac{1}{2}}}(d_0, 0) + 2p_{-\sqrt{\frac{1}{2}}}(d_0/\sqrt{5}, 0). \tag{23}$$

4. The characteristic equation of the matrix of the quadratic form $\sum_{i=1}^{n-2} (\Delta^2 x_i)^2$

The matrix A of the quadratic form

$$\sum_{i=1}^{n-2} (\Delta^2 x_i)^2 \equiv \sum_{i=1}^{n-2} (x_i - 2x_{i+1} + x_{i+2})^2 \equiv xAx' \tag{24}$$

is given by

$$A = \begin{pmatrix} 1 & -2 & 1 & . & . & . & . & . & . & . \\ -2 & 5 & -4 & 1 & . & . & . & . & . & . \\ 1 & -4 & 6 & -4 & 1 & . & . & . & . & . \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ . & . & . & . & . & 1 & -4 & 6 & -4 & 1 \\ . & . & . & . & . & . & 1 & -4 & 5 & -2 \\ . & . & . & . & . & . & . & 1 & -2 & 1 \end{pmatrix}, \tag{25}$$

and therefore the characteristic roots of A are given by

$$|A - \lambda I| = 0. \tag{26}$$

If $\{x_i\}$ is the characteristic vector corresponding to a root λ of (26) the following equations are satisfied:

$$\left.\begin{aligned} (1-\lambda)x_1 - 2x_2 + x_3 &= 0, \\ -2x_1 + (5-\lambda)x_2 - 4x_3 + x_4 &= 0, \\ x_{r-2} - 4x_{r-1} + (6-\lambda)x_r - 4x_{r+1} + x_{r+2} &= 0 \quad (r = 3, 4, \ldots, n-2), \\ x_{n-3} - 4x_{n-2} + (5-\lambda)x_{n-1} - 2x_n &= 0, \\ x_{n-2} - 2x_{n-1} + (1-\lambda)x_n &= 0. \end{aligned}\right\} \tag{27}$$

We have therefore to find the solution of the difference equation

$$x_{r-2} - 4x_{r-1} + (6-\lambda)x_r - 4x_{r+1} + x_{r+2} = 0 \quad (r = 1, 2, \ldots, n), \tag{28}$$

which satisfies the boundary conditions

$$\left.\begin{aligned} x_0 - 2x_1 + x_2 &= 0, \qquad x_{-1} - 2x_0 + x_1 = 0, \\ \text{and} \qquad x_{n-1} - 2x_n + x_{n+1} &= 0, \quad x_n - 2x_{n+1} + x_{n+2} = 0. \end{aligned}\right\} \tag{29}$$

Let $x_r = \rho^r$ be a solution of (28), then ρ satisfies the equation

$$\rho^2 + \rho^{-2} - 4(\rho + \rho^{-1}) + 6 - \lambda = 0. \tag{30}$$

Substituting $\rho + \rho^{-1} = u$ we get for u the quadratic $u^2 - 4u + 4 - \lambda = 0$ which has the roots $u = 2 \pm \lambda^{\frac{1}{2}}$. Since $\rho + \rho^{-1} = u$, i.e. $\rho^2 - \rho u + 1 = 0$, $\rho = \frac{1}{2}\{u \pm \sqrt{(u^2 - 4)}\}$, which leads to the following two pairs of roots of (30):

$$\left.\begin{aligned} \rho_1, \rho_2 &= \tfrac{1}{2}\{2 + \lambda^{\frac{1}{2}} \pm \sqrt{(\lambda + 4\lambda^{\frac{1}{2}})}\}, \\ \rho_3, \rho_4 &= \tfrac{1}{2}\{2 - \lambda^{\frac{1}{2}} \pm \sqrt{(\lambda - 4\lambda^{\frac{1}{2}})}\}. \end{aligned}\right\} \tag{31}$$

Now $\rho_1\rho_2 = \rho_3\rho_4 = 1$, and we may therefore take

$$\left.\begin{aligned} \rho_1 &= e^{\phi}, \quad \rho_2 = e^{-\phi}, \\ \rho_3 &= e^{i\theta}, \quad \rho_4 = e^{-i\theta}, \end{aligned}\right\} \tag{32}$$

where $$2 + \lambda^{\frac{1}{2}} = 2\cosh\phi \quad \text{and} \quad 2 - \lambda^{\frac{1}{2}} = 2\cos\theta, \tag{33}$$

and write the general solution of (28) as

$$x_r = A\cos r\theta + B\sin r\theta + C\cosh r\phi + D\sinh r\phi, \tag{34}$$

with the boundary conditions given by (29). If we substitute (34) in these boundary conditions and eliminate A, B, C, D we obtain an equation in θ and ϕ which is the characteristic equation provided we substitute for θ and ϕ in terms of λ from (33).

By altering the notation for the characteristic vector it is possible to simplify the characteristic equation and split it into two equations. Let us take, for instance, $n = 2m + 1$ and write the characteristic vector

$$\{x_i\} = \{x_{-m}, x_{-(m-1)}, \ldots, x_{-1}, x_0, x_1, \ldots, x_{m-1}, x_m\}.$$

Then the boundary conditions (29) becomes

$$\left.\begin{aligned} x_{-(m+1)} - 2x_{-m} + x_{-(m-1)} &= 0, \quad x_{-(m+2)} + 2x_{-(m+1)} + x_{-m} = 0, \\ x_{m+1} - 2x_m + x_{m-1} &= 0, \qquad x_{m+2} - 2x_{m+1} + x_m = 0. \end{aligned}\right\} \tag{35}$$

The substitution from (34) in (35) and the subsequent elimination of A, B, C, D gives

$$\begin{vmatrix} \cos m\theta, & \cosh m\phi, & \sin m\theta, & -\sinh m\phi, \\ \cos\overline{m+1}\,\theta, & \cosh\overline{m+1}\,\phi, & \sin\overline{m+1}\,\theta, & -\sinh\overline{m+1}\,\phi, \\ \cos m\theta, & \cosh m\phi, & -\sin m\theta, & \sinh m\phi, \\ \cos\overline{m+1}\,\theta, & \cosh\overline{m+1}\,\phi, & -\sin\overline{m+1}\,\theta, & \sinh\overline{m+1}\,\phi, \end{vmatrix} = 0$$

which splits into $$\cos m\theta \cosh\overline{m+1}\,\phi - \cos\overline{m+1}\,\theta \cosh m\phi = 0, \tag{36}$$

and $$\sin m\theta \sinh\overline{m+1}\,\phi - \sin\overline{m+1}\,\theta \sinh m\phi = 0. \tag{37}$$

It seems to be difficult to obtain the roots λ in the general case. But for any given value of m (i.e. of $n = 2m+1$) we can expand $\cos m\theta$, $\cos\overline{m+1}\,\phi$, etc.; substitute λ from (33) and obtain the two equations in λ whose roots are the characteristic roots of A. The case when n is even ($= 2m$) may be treated in a similar way by taking

$$\{x_i\} = \{x_{-(m+\frac{1}{2})}, x_{-(m-\frac{1}{2})}, \ldots, x_{-\frac{1}{2}}, x_{\frac{1}{2}}, \ldots, x_{m-\frac{1}{2}}, x_{m+\frac{1}{2}}\},$$

which leads us to the same equations as (36) and (37) with m replaced by $m-\frac{1}{2}$. To illustrate the procedure let $n = 5, 7, 9$, then the equations in λ are found to be:

$$n = 5: \quad \lambda^2 - 13\lambda + 10 = 0, \qquad \lambda - 5 = 0, \tag{38}$$

$$n = 7: \quad \lambda^3 - 19\lambda^2 + 70\lambda - 14 = 0, \qquad \lambda^2 - 11\lambda + 14 = 0, \tag{39}$$

$$n = 9: \quad \lambda^4 - 25\lambda^3 + 167\lambda^2 - 246\lambda + 18 = 0, \quad \lambda^3 - 17\lambda^2 + 63\lambda - 30 = 0. \tag{40}$$

It can be proved from (36) and (37) that each one of them gives one zero root for λ. The equations (38) to (40) therefore give the remaining $n-2 = 2m-1$ roots in each case.

The equations for the characteristic roots obtained above provide a useful check on the expressions for moments obtained by elaborate algebra in § 2·2 above. Let λ_i ($i = 1, 2, \ldots, n-2$) be the characteristic roots of A, then by an orthogonal transformation we can transform

$$\sum_{i=1}^{n-2} (\Delta^2 x_i)^2 \equiv xAx' = \sum_{i=1}^{n-2} \lambda_i z_i^2, \tag{41}$$

where z_i are independent normal variates with $\sigma(z_i) = \sigma(x_i)$. Taking $\sigma(x_i) = 1$, the characteristic function of $\Sigma(\Delta^2 x_i)^2$ can be expressed as

$$\begin{aligned} \phi(t) &= \frac{1}{(2\pi)^{\frac{1}{2}(n-2)}} \int_{-\infty}^{\infty} \ldots \int \exp\left[-\tfrac{1}{2}\Sigma z_r^2 + it\Sigma\lambda_r z_r^2\right] \pi\, dz_r \\ &= \prod_{r=1}^{n-2} (1 - 2i\lambda_r t)^{-\frac{1}{2}}. \end{aligned} \tag{42}$$

The semi-invariants or cumulants of $\Sigma(\Delta^2 x_i)^2$ can therefore be written down from $\phi(t)$. The first four moments can be shown to be

$$\left.\begin{aligned} \mu_1' &= \Sigma\lambda_i, \quad & \mu_2 &= 2\Sigma\lambda_i^2, \\ \mu_3 &= 8\Sigma\lambda_i^3, \quad & \mu_4 &= 3\mu_2^2 + 48\Sigma\lambda_i^4. \end{aligned}\right\} \tag{43}$$

For instance for $n = 9$, from (40) we have

$$\Sigma\lambda_i = 42, \quad \Sigma\lambda_i^2 = 454, \quad \Sigma\lambda_i^3 = 5628, \quad \Sigma\lambda_i^4 = 74102,$$

which confirm the formulae (8) obtained in § 2·2 above.

We may point out the similarity of the problem of solving the difference equation (28) with boundary conditions (29) with the problem of transverse vibrations in a rod with both ends clamped (see, for example, R. Courant & D. Hilbert, 1930, p. 253).

Finally, I wish to express my sincere thanks to Prof. E. S. Pearson and Dr H. O. Hartley for their advice in the preparation of this paper.

REFERENCES

ANDERSON, O. (1927). *J. R. Statist. Soc.* **90**, 548.
COURANT, R. & HILBERT, D. (1930). *Methoden der Mathimatischen Physik*. Berlin: Springer.
GUEST, P. G. (1951). *Suppl. J. R. Statist. Soc.* **13**, 233.
KAMAT, A. R. (1953*a*). *Biometrika*, **40**, 20.
KAMAT, A. R. (1953*b*). *Biometrika*, **40**, 116.
KAMAT, A. R. (1953*c*). *J. R. Statist. Soc.* (B), **15**, 233.
KEEN, J. & PAGE, D. J. (1953). *Appl. Statist.* **2**, 13.
MORSE, A. P. & GRUBBS, F. E. (1947). *Ann. Math. Statist.* **18**, 194.
PEARSON, K. (1931). *Tables for Statisticians and Biometricians*, **2**. London: Cambridge University Press for the Biometrika Trust.
TINTNER, G. (1940). *The Variate-Difference Method*. Indiana: Bloomington Press.
TINTNER, G. (1952). *Econometrics*. New York: John Wiley and Sons, Inc.
VON NEUMANN, J., KENT, R. H., BELLINSON, H. R. & HART, B. I. (1941). *Ann. Math. Statist.* **12**, 153.

THE STATISTICAL TREATMENT OF MEAN DEVIATION

By J. H. CADWELL
Ordnance Board

1. Introduction

In small samples it is easier to compute the mean deviation than to find the mathematically more tractable standard deviation. As a consequence, mean deviation is the generally used measure of dispersion in some types of work.

In the case of normal variation, the standard deviation, when corrected for bias, is the most efficient estimator of population standard deviation. However, in small samples the relative efficiency of mean deviation is high; for example, it is 91 % for samples of ten.

A more serious drawback of mean deviation is the rather troublesome form of its distribution for samples from a normal population. As a result of this, counterparts to methods available for standard deviations have not been developed. For instance, while two sets of variances can readily be compared because of the additive property of the χ^2 distribution and the existence of the F-ratio tables, the same problem for mean deviations has not been solved. A normal approximation to the distribution of mean deviation can be used for large sample sizes. It can also be used for the average of a number of independent values for a fixed small sample size. However, this still leaves many situations where it is inapplicable.

It is the object of this note to give approximate methods applicable to some frequently needed procedures. The errors involved are not much greater than those caused by rounding the sample values of mean deviation to three significant figures. They should be tolerable in most cases.

In §§ 2 and 3 the basis of the method is explained, and three applications are discussed in subsequent sections.

2. The basic approximation

In a previous paper (Cadwell, 1953) the use of a power of χ^2 as an approximation to the distribution of mean deviation is considered. This gives a result of the form

$$m/\sigma \text{ is approximately distributed like } (\chi^2/c)^\alpha.$$

The variable power used, together with the degrees of freedom of χ^2 and the multiplicative constant, allow the first three moments of the two distributions to be matched.

The dependence of the power to be used on sample size restricts the method to sets of values of mean deviation with a common sample size. By choosing a suitable fixed power we can ensure agreement of the first two moments with a small discrepancy for the third moment. The ability to work with sets of varying sample size adds greatly to the flexibility of the approximation.

We use $\overline{m}(k, n)$ to denote the average of k values of mean deviation, each for a sample of size n from a normal population of standard deviation σ. We find that

$$c\left\{\frac{\overline{m}(k,n)}{\sigma}\right\}^{1\cdot8} \tag{1}$$

has approximately the χ^2 distribution on ν degrees of freedom. The values of c and ν are given by

$$\nu = \nu_0 + 0{\cdot}196 - \frac{0{\cdot}159}{\nu_0} + \ldots, \quad \text{where} \quad \nu_0 = \frac{0{\cdot}617k}{V_n^2};$$

$$\log c = \log 2 + 1{\cdot}8\{\log \Gamma(0{\cdot}\dot{5} + \tfrac{1}{2}\nu) - \log \Gamma(\tfrac{1}{2}\nu) - \log \mathscr{E}(\overline{m}_n/\sigma)\}.$$

Here $\mathscr{E}(\overline{m}_n/\sigma)$ and V_n are respectively the mean value and coefficient of variation of the mean deviation in a sample of n. The value of ν can be rounded to 0·1 unit with little loss of accuracy. It is advisable to retain three or four figures in c.

Table 1 gives values of c and ν for n from 4 to 10 and k from 1 to 10. Table 2 gives values for n from 10 to 50 by steps of 5, and for k from 1 to 5. The value of ν is found by linear interpolation for n, and then rounded to 0·1 (or to an integer for values of ν greater than 20). The constant c is determined by linear interpolation for this value of ν, i.e. c is regarded as a function of ν and not of n.

3. Accuracy of the approximation

The power of χ^2 required for a three-moment fit runs from 0·5 when $n = 2$ to 0·594 when $n = \infty$. Empirical work shows that a single value of 0·$\dot{5}$ (reciprocal 1·8) gives a reasonable compromise for values of n greater than 3 (see Table 1 of my previous paper (Cadwell, 1953)). By using Godwin & Hartley's tables (1945) we can examine the accuracy of the approximation (1) up to $n = 10$. Beyond this, the more refined method of my previous paper gives a measure of error due to the present approximation. Using (1) and tables of the χ^2 integral we evaluate

$$\Pr\{\text{mean deviation} < m_0\sigma\}$$

approximately and compare the result with the true value.

The error is found to decrease for sample sizes from 4 up to about 7. It then increases again up to sample size 15, thereafter tending steadily to zero. Errors are shown below for $n = 5, 10$:

$n = 5$	m_0	0·17	0·32	0·49	0·70	0·93	1·19	1·51
	True value	0·0050	0·0527	0·2100	0·5122	0·7978	0·9510	0·9951
	Error	−0·0002	−0·0003	+0·0006	+0·0003	−0·0004	0·0000	+0·0001
$n = 10$	m_0	0·33	0·46	0·59	0·75	0·91	1·09	1·30
	True value	0·0046	0·0474	0·1933	0·5074	0·7950	0·9520	0·9951
	Error	+0·0003	+0·0010	+0·0006	−0·0013	−0·0011	0·0000	+0·0001

When k exceeds 1 no exact values are available; the following procedure was therefore adopted.

Using series expansions for β_1 and β_2 due to Geary, and given in an Editorial Note to Godwin & Hartley's paper (1945), we find the β_1 and β_2 differences of the distributions of

$$(\chi^2/c)^{0{\cdot}\dot{5}} \quad \text{and} \quad \overline{m}(k, n).$$

Results are shown below for $n = 5$, 10 and $k = 1$ to 4:

n	k	$\beta_1(\overline{m})$	$\beta_2(\overline{m})$	β_1 difference	β_2 difference
5	1	0·230	3·197	−0·002	−0·020
	2	0·115	3·098	−0·007	−0·029
	3	0·077	3·066	−0·006	−0·026
	4	0·057	3·049	−0·005	−0·019
10	1	0·106	3·093	−0·009	−0·034
	2	0·053	3·046	−0·006	−0·019
	3	0·035	3·031	−0·004	−0·013
	4	0·026	3·023	−0·003	−0·009

Apart from an increase from $k = 1$ to $k = 2$ when $n = 5$, differences in the β's decrease as k increases. It will be seen that $n = 5$, $k = 2$ and $n = 10$, $k = 1$ have roughly the same β values and give nearly the same β differences. Thus errors in the first case should be similar to those in the second and these have been examined above.

4. Tests of homogeneity

Given a set of k mean deviations supposedly for samples from normal populations of a fixed standard deviation, we may suspect that the variability present is greater than can be attributed to sampling fluctuations. In order to test this point we use (1) to transform the set of mean deviations to a fixed multiple of approximate χ^2 values on $\nu_1, \nu_2, \ldots, \nu_k$ degrees of freedom respectively. We can now apply the Neyman-Pearson L test with Bartlett's modification to test the homogeneity of these χ^2 values.

When all samples are of the same size n, a rapid but somewhat less powerful test is available. We evaluate the ratio of maximum and minimum values in the set. If this ratio exceeds the value given in Table 5 of my previous paper (Cadwell, 1953, p. 346), we assume that some sort of heterogeneity is present. This procedure will result in an incorrect decision in approximately 5 % of cases when the values are homogeneous. Table 6 of the same paper gives the corresponding 1 % points for this ratio. The exact probability will seldom differ from the nominal value by as much as 0·2 % when using these tables.

5. Estimation of standard deviation

Suppose we have a set of values $m_1, m_2, \ldots, m_k$ from populations of fixed dispersion. Using the additive property of χ^2, we shall have:

$$\Sigma c_i (m_i/\sigma)^{1\cdot 8} \text{ has approximately the } \chi^2 \text{ distribution on } \Sigma\nu_i \text{ d.f.},$$

where c_i and ν_i are obtained from the first column of Table 1 to correspond to the sample size of the ith mean deviation. This will furnish the estimate

$$\hat{\sigma} = \left(\frac{\Sigma c_i m_i^{1\cdot 8}}{\Sigma \nu_i}\right)^{0\cdot 5}.$$

Confidence limits for σ can readily be found from tables of the χ^2 integral. Thus the true value will be less than

$$\left(\frac{\Sigma c_i m_i^{1\cdot 8}}{\chi^2_{0\cdot 05}}\right)^{0\cdot 5}$$

on 95 % of occasions. If we require limits that include the true value on 95 % of occasions, we shall take

$$\left\{\frac{\Sigma c_i m_i^{1\cdot8}}{\chi^2_{0\cdot975}}\right\}^{0\cdot5} \quad \text{and} \quad \left\{\frac{\Sigma c_i m_i^{1\cdot8}}{\chi^2_{0\cdot025}}\right\}^{0\cdot5}.$$

When all k samples have the same size n, we can use the alternative estimate

$$\hat{\sigma} = \frac{\overline{m}(k,n)}{\mathscr{E}(\overline{m}_n/\sigma)}, \quad \text{where} \quad \mathscr{E}(\overline{m}_n/\sigma) = \sqrt{\frac{2(n-1)}{n\pi}}.$$

Values of $\mathscr{E}(\overline{m}_n/\sigma)$ are given in Tables 1 and 2. Tables 3 (*a*)–(*d*) enable either one- or two-sided 95 % limits to be found. Thus on 95 % of occasions the true value of σ will lie between

$$\overline{m}(k,n)/\overline{m}_{0\cdot975}(k,n) \quad \text{and} \quad \overline{m}(k,n)/\overline{m}_{0\cdot025}(k,n).$$

Errors in Table 3 will seldom exceed two units in the third decimal.

Should it be necessary to deal with values of k beyond 10 a normal approximation to the distribution of $\overline{m}(k,n)$ can be used; errors in the 2·5 % levels will not exceed 0·01 with n as small as 5.

The expected value is given above and the variance is given by

$$\operatorname{var}\frac{\overline{m}(k,n)}{\sigma} = \frac{1}{k}\operatorname{var}\frac{\overline{m}(1,n)}{\sigma} = \frac{2(n-1)}{kn^2\pi}\left\{\tfrac{1}{2}\pi + \sqrt{\{n(n-2)\}} - n + \sin^{-1}\frac{1}{n-1}\right\}.$$

Values of $\operatorname{var}\overline{m}(1,n)/\sigma$ are given in Tables 1 and 2.

6. Comparison of two sets of mean deviations

It is often necessary to compare two sets of samples to see if the dispersion can be regarded as the same in the two cases.

On the hypothesis of a common variance in the two sets, we shall have

$$\frac{\Sigma c_i m_i^{1\cdot8}\Sigma\nu'_i}{\Sigma c'_i (m'_i)^{1\cdot8}\Sigma\nu_i}$$

is approximately an F-ratio on $\Sigma\nu_i$ and $\Sigma\nu'_i$ degrees of freedom. Working at the 5 % level we shall reject the hypothesis of a common variance if either this ratio or its reciprocal exceeds the upper 2·5 % level of F on the appropriate numbers of degrees of freedom.

When the sample sizes are constant within the two sets, we have a simpler test based on

$$\frac{c\nu'}{c'\nu}\left\{\frac{\overline{m}(k,n)}{\overline{m}(k',n')}\right\}^{1\cdot8}.$$

The values of c and ν are obtained from Table 1 entered with n and k and similarly for c' and ν'. This ratio and its reciprocal will be compared with the upper 2·5 % point of F on ν and ν' degrees of freedom.

It will be seen that fractional degrees of freedom are involved. However, in the majority of cases significance or non-significance can be settled without elaborate interpolation.

Where it can be used, the latter approximation is to be preferred. It involves two cases of (1) for values of k usually greater than unity, while the other depends on a number of applications of (1) each with $k = 1$. As the errors in (1) fall rapidly with increase of k, the second approximation should be more accurate as well as being simpler to apply. The degrees

of freedom involved in the second method are a little less than in the first. However, the difference in power of the two tests should be negligible. The same argument will apply to the similar choice of methods that may arise in § 5.

I should like to thank Mr D. F. Mills for his help in the preparation of the tables. Acknowledgement is made to the Chief Scientist, Ministry of Supply, for permission to publish this paper.

REFERENCES

CADWELL, J. H. (1953). *Biometrika*, **40**, 336.
GODWIN, H. J. & HARTLEY, H. O. (1945). *Biometrika*, **33**, 210.

Table 1. *Transformation constants, c and ν. The upper figure (in bold type) is that of ν and the lower is c*

n \ k	1	2	3	4	5	6	7	8	9	10	$\mathscr{E}\left(\frac{\overline{m}_n}{\sigma}\right)$	$\text{var}\,\frac{\overline{m}(1,n)}{\sigma}$
4	**3·5**	**6·9**	**10·2**	**13·6**	**16·9**	**20·2**	**23·6**	**26·9**	**30·3**	**33·6**		
	6·013	12·59	19·00	25·61	32·02	38·44	45·05	51·47	58·08	64·50	0·6910	0·08818
5	**4·6**	**9·0**	**13·5**	**17·9**	**22·3**	**26·8**	**31·2**	**35·6**	**40·1**	**44·5**		
	7·677	15·73	23·98	32·05	40·12	48·38	56·46	64·53	72·79	80·86	0·7136	0·07094
6	**5·7**	**11·2**	**16·7**	**22·3**	**27·8**	**33·3**	**38·8**	**44·3**	**49·9**	**55·4**		
	9·336	19·05	28·77	38·67	48·40	58·13	67·86	77·59	87·50	97·23	0·7284	0·05934
7	**6·8**	**13·4**	**20·0**	**26·6**	**33·2**	**39·8**	**46·4**	**53·0**	**59·6**	**66·2**		
	10·99	22·36	33·74	45·12	56·51	67·89	79·28	90·66	102·0	113·4	0·7387	0·05101
8	**7·9**	**15·6**	**23·3**	**30·9**	**38·6**	**46·3**	**54·0**	**61·7**	**69·4**	**77·1**		
	12·65	25·67	38·71	51·57	64·61	77·65	90·68	103·7	116·8	129·8	0·7464	0·04473
9	**9·0**	**17·7**	**26·5**	**35·3**	**44·1**	**52·8**	**61·6**	**70·4**	**79·1**	**87·9**		
	14·30	28·82	43·50	58·19	72·88	87·40	102·1	116·8	131·3	146·0	0·7522	0·03982
10	**10·0**	**19·9**	**29·8**	**39·6**	**49·5**	**59·3**	**69·2**	**79·0**	**88·9**	**98·7**		
	15·79	32·12	48·47	64·64	80·98	97·16	113·5	129·7	146·0	162·2	0·7569	0·03589

Table 2. *Transformation constants, c and ν*

n \ k	1		2		3		4		5		$\mathscr{E}\left(\frac{\overline{m}_n}{\sigma}\right)$	$\text{var}\,\frac{\overline{m}(1,n)}{\sigma}$
	ν	c	ν	c	ν	c	ν	c	ν	c		
10	10·0	15·79	19·9	32·12	29·8	48·47	39·6	64·64	49·5	80·98	0·7569	0·03589
15	15·5	24·07	30·7	48·34	46·0	72·78	61·2	97·07	76·5	121·5	0·7708	0·02403
20	20·9	32·17	41·5	64·56	62·2	97·11	82·9	129·7	103·5	162·0	0·7777	0·01806
25	26·3	40·28	52·3	80·78	78·4	121·4	104·5	162·1	130·5	202·6	0·7813	0·01446
30	31·7	48·39	63·2	97·15	94·6	145·8	126·1	194·5	157·6	243·3	0·7845	0·01206
35	37·1	56·50	74·0	113·4	110·8	170·1	147·7	226·9	184·6	283·8	0·7864	0·01035
40	42·5	64·60	84·8	129·6	127·1	194·5	169·4	259·5	211·6	324·3	0·7878	0·00906
45	47·9	72·71	95·6	145·8	143·3	218·9	191·0	292·0	238·6	364·9	0·7890	0·00805
50	53·3	80·82	106·4	162·0	159·5	243·2	212·6	324·4	265·7	405·6	0·7899	0·00725

The symbol $\overline{m}(k, n)$ denotes the average of k independent mean deviations for samples of size n from normal populations of standard deviation σ.

In Tables 1 and 2 the last two columns give the expected value of $\overline{m}(k, n)/\sigma$ and the variance of $\overline{m}(1, n)/\sigma$ respectively. In addition we have the result:

$c\{\overline{m}(k, n)/\sigma\}^{1\cdot 8}$ has approximately the χ^2 distribution on ν d.f.

Table 3. *Percentage points of $\overline{m}(k, n)$, the averages of k independent mean deviations, each for a sample of n from a normal population of unit standard deviation*

(*a*) Lower 2·5 % points

n \ k	1	2	3	4	5	6	7	8	9	10
4	0·199	0·323	0·382	0·420	0·446	0·466	0·482	0·495	0·505	0·514
5	0·260	0·376	0·433	0·468	0·492	0·510	0·525	0·536	0·546	0·554
6	0·306	0·416	0·469	0·502	0·525	0·542	0·555	0·566	0·575	0·582
7	0·342	0·447	0·497	0·528	0·549	0·565	0·577	0·587	0·596	0·603
8	0·372	0·472	0·519	0·548	0·568	0·583	0·595	0·604	0·612	0·619
9	0·396	0·492	0·537	0·565	0·584	0·598	0·609	0·618	0·625	0·632
10	0·417	0·509	0·552	0·578	0·597	0·610	0·621	0·629	0·636	0·642

(*b*) Lower 5 % points

n \ k	1	2	3	4	5	6	7	8	9	10
4	0·254	0·371	0·425	0·459	0·482	0·499	0·513	0·524	0·533	0·541
5	0·315	0·422	0·473	0·503	0·525	0·541	0·553	0·563	0·572	0·579
6	0·360	0·460	0·507	0·535	0·555	0·570	0·581	0·590	0·598	0·604
7	0·394	0·489	0·532	0·559	0·576	0·591	0·601	0·610	0·618	0·624
8	0·422	0·512	0·553	0·578	0·595	0·608	0·618	0·626	0·633	0·638
9	0·445	0·529	0·569	0·593	0·609	0·621	0·631	0·638	0·645	0·650
10	0·464	0·545	0·583	0·605	0·621	0·633	0·642	0·649	0·655	0·660

(*c*) Upper 5 % points

n \ k	1	2	3	4	5	6	7	8	9	10
4	1·224	1·057	0·988	0·945	0·918	0·898	0·882	0·869	0·858	0·850
5	1·187	1·041	0·978	0·941	0·916	0·898	0·884	0·873	0·863	0·856
6	1·158	1·026	0·969	0·935	0·913	0·897	0·884	0·874	0·865	0·858
7	1·135	1·013	0·961	0·930	0·910	0·894	0·882	0·873	0·865	0·859
8	1·116	1·002	0·954	0·925	0·906	0·892	0·881	0·872	0·864	0·858
9	1·100	0·994	0·948	0·921	0·903	0·889	0·879	0·870	0·864	0·858
10	1·086	0·985	0·942	0·917	0·899	0·887	0·877	0·869	0·863	0·857

(*d*) Upper 2·5 % points

n \ k	1	2	3	4	5	6	7	8	9	10
4	1·344	1·137	1·051	0·999	0·966	0·941	0·921	0·906	0·893	0·882
5	1·292	1·111	1·033	0·989	0·958	0·936	0·919	0·905	0·894	0·884
6	1·253	1·089	1·020	0·978	0·951	0·931	0·915	0·903	0·892	0·884
7	1·222	1·071	1·007	0·970	0·944	0·926	0·911	0·900	0·891	0·883
8	1·196	1·056	0·996	0·962	0·938	0·921	0·908	0·897	0·888	0·881
9	1·175	1·044	0·988	0·955	0·933	0·917	0·904	0·894	0·886	0·879
10	1·156	1·033	0·980	0·949	0·928	0·913	0·901	0·891	0·884	0·877

The values in Tables 3(*a*)–(*d*) are exact for $k = 1$. For other values of k, the error will not exceed 0·003 and will usually be much smaller.

TESTS OF LINEAR HYPOTHESES IN UNIVARIATE AND MULTIVARIATE ANALYSIS WHEN THE RATIOS OF THE POPULATION VARIANCES ARE UNKNOWN

By G. S. JAMES
University of Leeds

1. Introduction

In a previous paper (James, 1951) I have attempted to generalize B. L. Welch's method of comparing two means when the ratio of the population variances is unknown (Welch, 1947*a*, *b*; Aspin, 1948, 1949) to the case where one requires a test for the equality of several means. Welch (1951) has derived a similar test which makes use of tables of the variance ratio instead of those of χ^2, and has shown that the two tests are equivalent to the first order of inverse powers of the sample sizes, which is the order to which he worked. These results can easily be applied to certain other problems, such as comparison of the slopes of regression lines, but these extensions are, from the mathematical point of view, trivial, for they merely depend on our ability to calculate from the data quantities which are distributed in forms to which the original arguments can be applied.

The essential difference between Welch's original problem (considered as a test of significance) and my own was that the linear hypothesis to be tested imposed one constraint in his case and several in mine. However, the problem of several means is a very special one, so in § 2 of the present paper we consider a more general set-up. § 3 consists of simple illustrations of this general work. In § 4 applications are made to a simple factorial experimental lay-out, in which no assumption about constancy of the error variances can be made, but in which they can be individually estimated from replicate determinations. The work of § 2 is generalized to the multivariate case in § 5, and in §§ 6 and 7 the results are applied to the multivariate problems of two and several samples, respectively. The two-sample problem is the multivariate analogue of the Behrens-Fisher problem, as treated by Welch. § 8 consists of numerical illustrations of the multivariate tests of §§ 6 and 7.

The reader's attention is drawn to the remarks at the end of § 4; it is hoped to investigate the problem further along the lines suggested there, and to publish any reasonably simple techniques found for dealing with the sort of problem which has been encountered in practice. Meanwhile, it will be noted that § 4 is not essential to the understanding of the rest of the paper.

2. A univariate linear hypothesis

Let $x_1, \ldots, x_n$ be n random variables whose mean values are known linear functions of k unknown parameters $\theta_1, \ldots, \theta_k$, and which are distributed independently and normally about these mean values with unknown variances $\alpha_1, \ldots, \alpha_n$. In matrix notation we may write

$$\mathscr{E}(\mathbf{x}) = \mathbf{B}\boldsymbol{\theta}, \tag{2·1}$$

where $\mathbf{x}$ and $\boldsymbol{\theta}$ are the column vectors $\{x_1, \ldots, x_n\}$ and $\{\theta_1, \ldots, \theta_k\}$, and $\mathbf{B} = (b_{ij})$ is an $n \times k$ matrix of known constants.† (We assume that the rank of $\mathbf{B}$ is k.) Let a_i $(i = 1, \ldots, n)$ be

† Any term independent of the θ_j which would appear on the right-hand side of (2·1) can be incorporated into $\mathbf{x}$ by a transformation.

unbiased estimates of the corresponding α_i, based on ν_i degrees of freedom, and distributed independently of the x_i and of each other in the usual χ^2-type distribution.

We now suppose that the null hypothesis to be tested is that $\theta_1 = \ldots = \theta_r = 0$ (r being some number between 1 and k), the alternative hypothesis being that the whole set of parameters $\theta_1, \ldots, \theta_k$ is entirely unrestricted. Often an apparently more general linear hypothesis occurs, in which the parameters have constraints imposed upon them even under the alternative hypothesis, and in which the null hypothesis states that the parameters satisfy a further set of r linearly independent linear relations; but this type of hypothesis may always be reduced to the above canonical form by means of suitable linear transformations of the parameters. (See, for example, Wilks (1946, pp. 171–3); examples of the reduction occur in later sections of this paper.) If we partition $\boldsymbol{\theta}$ according to its first r and last $(k-r)$ elements, thus: $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$, then the above null hypothesis may be written as $\boldsymbol{\theta}_1 = 0$.

Now suppose for a moment that the α_i were known. Then, under the alternative hypothesis, the value of $\boldsymbol{\theta}$ which maximized the likelihood of the observed $\mathbf{x}$ would be

$$\hat{\boldsymbol{\theta}} = \mathbf{C}^{-1}\mathbf{d}, \tag{2·2}$$

where

$$\mathbf{C} = \mathbf{B}'\boldsymbol{\alpha}^{-1}\mathbf{B}, \quad \mathbf{d} = \mathbf{B}'\boldsymbol{\alpha}^{-1}\mathbf{x}, \quad \boldsymbol{\alpha} = \operatorname{diag}(\alpha_1, \ldots, \alpha_n). \tag{2·3}$$

Also the variance matrix of the estimator $\hat{\boldsymbol{\theta}}$ would be

$$\operatorname{var}\hat{\boldsymbol{\theta}} = \mathbf{C}^{-1}. \tag{2·4}$$

Moreover, it is known that, under these circumstances, the test for the null hypothesis derived by the likelihood-ratio method would depend on the fact that, when this hypothesis is true,

$$\hat{\boldsymbol{\theta}}_1'\mathbf{V}^{-1}\hat{\boldsymbol{\theta}}_1 \text{ is distributed as } \chi^2 \text{ on } r \text{ d.f.}, \tag{2·5}$$

where $\mathbf{V}$ is the variance matrix of $\hat{\boldsymbol{\theta}}_1$ (namely, the first r rows and columns of $\mathbf{C}^{-1}$). That is to say, writing $\mathbf{t}$ in place of $\hat{\boldsymbol{\theta}}_1$ for brevity, and using the notation $\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha)$ as a reminder that both $\mathbf{t}$ and $\mathbf{V}$ are, in general, functions of the α_i,

$$\Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha) \leqslant 2\xi] = G_\rho(\xi), \tag{2·6}$$

where 2ξ denotes the tabled value of χ^2 for a particular probability or significance level, $\rho = \frac{1}{2}r$ and

$$G_\rho(\xi) = [\Gamma(\rho)]^{-1}\int_0^\xi t^{\rho-1}e^{-t}\,dt. \tag{2·7}$$

(The symbols ξ and ρ are used in place of χ^2 and r, and the formula (2·6) written in terms of ξ rather than in terms of the corresponding probability or significance level, merely as a matter of later mathematical convenience. The final results will be translated back into familiar form.)

As the α_i are not in fact known, we may, if their estimates a_i are based on large numbers of degrees of freedom, use as an approximation the result

$$\Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(a) \leqslant 2\xi] = G_\rho(\xi), \tag{2·8}$$

where $\mathbf{t}(a)$ and $\mathbf{V}(a)$ are the same functions of the a_i as $\mathbf{t}(\alpha)$ and $\mathbf{V}(\alpha)$ are of the α_i. This suggests that in the general case we try to find a function $h(a)$ of the a_i (and of ξ) such that

$$\Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(a) \leqslant 2h(a)] = G_\rho(\xi). \tag{2·9}$$

In large samples $2h(a)$ will approach $2\xi \equiv \chi^2$, and we shall hope, as in previous work, to be able to write $h(a)$ as a series of terms of decreasing order of magnitude, as in (2·13) below.

Carrying out two successive symbolic Taylor expansions, (2·9) leads in the usual manner (see James (1951) or Welch (1947*a*)) first to

$$\Theta \Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha) \leqslant 2h(\alpha)] = G_\rho(\xi), \tag{2·10}$$

and then to

$$\Theta \exp[h_+(\alpha)\, D] \Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha) \leqslant 2\xi] = G_\rho(\xi), \tag{2·11}$$

where

$$\Theta = \Pi[e^{-\alpha_i \partial_i}(1-2\alpha_i\partial_i/\nu_i)^{-\frac{1}{2}\nu_i}]$$

$$= 1 + \Sigma \frac{\alpha_i^2 \partial_i^2}{\nu_i} + \left\{ \tfrac{4}{3} \Sigma \frac{\alpha_i^3 \partial_i^3}{\nu_i^2} + \frac{1}{2} \left(\Sigma \frac{\alpha_i^2 \partial_i^2}{\nu_i} \right)^2 \right\} + O(\nu^{-3}), \tag{2·12}$$

$$h(a) = \xi + h_+(a)$$

$$= \xi + h_1(a) + h_2(a) + \ldots, \tag{2·13}$$

($h_s(a)$ being of order ν^{-s}), and where ∂_i denotes $\partial/\partial\alpha_i$ and D denotes $\partial/\partial\xi$. Substituting these expressions for Θ and $h_+(a)$ into (2·11) and equating terms of successive orders, we obtain, exactly as before,

$$\left[h_1(\alpha)\, D + \Sigma \frac{\alpha_i^2 \partial_i^2}{\nu_i} \right] \Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha) \leqslant 2\xi] = 0, \tag{2·14}$$

$$\left[h_2(\alpha)\, D + \tfrac{1}{2} h_1^2(\alpha)\, D^2 + \Sigma \frac{\alpha_i^2}{\nu_i} (h_1^{(ii)}(\alpha)\, D + 2h_1^{(i)}(\alpha)\, \partial_i D + h_1(\alpha)\, \partial_i^2 D) \right.$$

$$\left. + \tfrac{4}{3} \Sigma \frac{\alpha_i^3 \partial_i^3}{\nu_i^2} + \frac{1}{2} \left(\Sigma \frac{\alpha_i^2 \partial_i^2}{\nu_i} \right)^2 \right] \Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha) \leqslant 2\xi] = 0, \tag{2·15}$$

as equations giving $h_1(\alpha)$, $h_2(\alpha)$, and hence $h_1(a)$, $h_2(a)$. Here $h_1^{(i)}(\alpha)$ denotes $\partial_i h_1(\alpha)$, $h_1^{(ij)}(\alpha)$ denotes $\partial_i \partial_j h_1(\alpha)$, and so on.

We employ the method used previously to evaluate derivatives like $\partial_i^3 \Pr[\ldots]$. Let

$$J = \Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha+\epsilon) \leqslant 2\xi]. \tag{2·16}$$

Then, by Taylor's theorem,

$$J = [1 + \Sigma\epsilon_i\partial_i + \tfrac{1}{2}\Sigma\epsilon_i\epsilon_j\partial_i\partial_j + \ldots] \Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha) \leqslant 2\xi]. \tag{2·17}$$

We also have, by definition,

$$J = (2\pi)^{-\frac{1}{2}r} |\mathbf{var}\,\mathbf{t}(\alpha+\epsilon)|^{-\frac{1}{2}} \int_{\tau'\mathbf{V}^{-1}(\alpha+\epsilon)\tau \leqslant 2\xi} \exp[-\tfrac{1}{2}\boldsymbol{\tau}'(\mathbf{var}\,\mathbf{t}(\alpha+\epsilon))^{-1}\boldsymbol{\tau}]\, d\tau, \tag{2·18}$$

where $\int \ldots d\tau$ denotes integration over the region indicated in the space of $\tau_1, \ldots, \tau_r$ (i.e. the space of realizable values of the random variables $t_1, \ldots, t_r$). It should be noted that $\mathbf{var}\,\mathbf{t}(\alpha+\epsilon)$ is not the same thing as $\mathbf{V}(\alpha+\epsilon)$, which is $\mathbf{var}\,\mathbf{t}(\alpha)$ with $\alpha_i+\epsilon_i$ substituted for the α_i; for the original random variables x_i, of which $\mathbf{t}(\alpha)$ is a function, are still supposed to have variances α_i, not $\alpha_i+\epsilon_i$. Now in (2·18) we make a real, non-singular linear transformation

$$\boldsymbol{\tau} = \mathbf{T}\mathbf{u} \tag{2·19}$$

to new variables u_i, which is such that

$$\tfrac{1}{2}\mathbf{T}'\mathbf{V}^{-1}(\alpha+\epsilon)\,\mathbf{T} = \mathbf{I}, \tag{2·20}$$

$$\tfrac{1}{2}\mathbf{T}'(\mathbf{var}\,\mathbf{t}(\alpha+\epsilon))^{-1}\,\mathbf{T} = \mathbf{I} - \boldsymbol{\eta}, \tag{2·21}$$

where $\mathbf{I}$ is the unit matrix and $\boldsymbol{\eta}$ is a diagonal matrix:

$$\boldsymbol{\eta} = \mathrm{diag}\,(\eta_1, \ldots, \eta_r). \tag{2·22}$$

$\mathbf{T}$ will be a function of the α_i and ϵ_i. Under this transformation (2·18) becomes

$$J = \pi^{-\frac{1}{2}r} |\mathbf{I}-\boldsymbol{\eta}|^{\frac{1}{2}} \int_{\mathbf{u}'\mathbf{u}\leqslant\xi} \exp[-\mathbf{u}'(\mathbf{I}-\boldsymbol{\eta})\mathbf{u}]\,du. \tag{2·23}$$

We now expand the factor $\exp[\mathbf{u}'\boldsymbol{\eta}\mathbf{u}] \equiv \exp[\Sigma\eta_i u_i^2]$ appearing in the integrand in powers of the η_i and integrate term by term. The individual integrals are of Dirichlet's type, and we obtain, exactly as in the paper previously cited (equation (26)),†

$$J = \left(\frac{|\mathbf{I}-\boldsymbol{\eta}E|}{|\mathbf{I}-\boldsymbol{\eta}|}\right)^{-\frac{1}{2}} G_\rho(\xi), \tag{2·24}$$

where we again use the operational notations $EG_\rho(\xi) = G_{\rho+1}(\xi)$, $\Delta = E-1$. Now, using equations (2·20) and (2·21), we have

$$\begin{aligned}\frac{|\mathbf{I}-\boldsymbol{\eta}E|}{|\mathbf{I}-\boldsymbol{\eta}|} &= \frac{|\mathbf{I}-\boldsymbol{\eta}-\boldsymbol{\eta}\Delta|}{|\mathbf{I}-\boldsymbol{\eta}|} \\ &= \frac{|\mathbf{var}^{-1}\mathbf{t}(\alpha+\epsilon)-[\mathbf{V}^{-1}(\alpha+\epsilon)-\mathbf{var}^{-1}\mathbf{t}(\alpha+\epsilon)]\Delta|}{|\mathbf{var}^{-1}\mathbf{t}(\alpha+\epsilon)|} \\ &= |\mathbf{I}-[\mathbf{V}^{-1}(\alpha+\epsilon).\mathbf{var}\,\mathbf{t}(\alpha+\epsilon)-\mathbf{I}]\Delta|. \end{aligned} \tag{2·25}$$

Thus, from (2·24),
$$J = |\mathbf{I}-\mathbf{X}\Delta|^{-\frac{1}{2}} G_\rho(\xi), \tag{2·26}$$

where
$$\mathbf{X} = \mathbf{V}^{-1}(\alpha+\epsilon).\mathbf{var}\,\mathbf{t}(\alpha+\epsilon)-\mathbf{I}. \tag{2·27}$$

Such expressions as $|\mathbf{I}-\mathbf{X}\Delta|^{-\frac{1}{2}}$ are not easily expanded directly in a form suitable to our needs, but we do have the special result (an analogue of the ordinary logarithmic series)

$$-\log|\mathbf{I}-\mathbf{Y}| = \operatorname{tr}\mathbf{Y} + \tfrac{1}{2}\operatorname{tr}\mathbf{Y}^2 + \tfrac{1}{3}\operatorname{tr}\mathbf{Y}^3 + \ldots. \tag{2·28}$$

If $\mathbf{Y}$ is a numerical matrix this series is convergent if all the eigenvalues of $\mathbf{Y}$ are numerically less than unity; a proof of this result will be presented in § 5. Using it in (2·26) we have

$$\begin{aligned} J &= \exp(-\tfrac{1}{2}\log|\mathbf{I}-\mathbf{X}\Delta|)\,G_\rho(\xi) \\ &= \exp(\tfrac{1}{2}\operatorname{tr}\mathbf{X}.\Delta + \tfrac{1}{4}\operatorname{tr}\mathbf{X}^2.\Delta^2+\ldots)\,G_\rho(\xi) \\ &= [1+\tfrac{1}{2}\operatorname{tr}\mathbf{X}.\Delta + \{\tfrac{1}{4}\operatorname{tr}\mathbf{X}^2+\tfrac{1}{8}(\operatorname{tr}\mathbf{X})^2\}\Delta^2+\ldots]\,G_\rho(\xi). \end{aligned} \tag{2·29}$$

We shall first of all treat the *simple case* in which $\mathbf{t}(\alpha)$ is actually independent of the α_i (as in the k means problem previously considered), deferring consideration of the more complicated *general case* until later. Since $\mathbf{t}(\alpha)$ is independent of the α_i, it follows firstly that $\mathbf{var}\,\mathbf{t}(\alpha+\epsilon) = \mathbf{var}\,\mathbf{t}(\alpha) = \mathbf{V}(\alpha)$, and secondly that $\mathbf{V}(\alpha)$ is a homogeneous linear function of the α_i (since $\mathbf{t}$ is a linear function of the x_i). Thus (2·27) becomes

$$\begin{aligned} \mathbf{X} &= \mathbf{V}^{-1}(\alpha+\epsilon)\,\mathbf{V}(\alpha)-\mathbf{I} \\ &= [\mathbf{V}(\alpha)+\Sigma\epsilon_i\mathbf{V}_i(\alpha)]^{-1}\,\mathbf{V}(\alpha)-\mathbf{I} \\ &= [\mathbf{I}+\Sigma\epsilon_i\mathbf{V}^{-1}\mathbf{V}_i]^{-1}-\mathbf{I} \\ &= -\Sigma\epsilon_i\mathbf{V}^{-1}\mathbf{V}_i + \Sigma\epsilon_i\epsilon_j\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j - \ldots, \end{aligned} \tag{2·30}$$

† In the second line of this equation a summation sign (referring to the b_i) was accidentally omitted.

where we abbreviate $\mathbf{V}(\alpha)$ to $\mathbf{V}$ and use the notation

$$\mathbf{V}_i = \partial \mathbf{V}/\partial\alpha_i, \quad \mathbf{V}_{i,j} = \partial^2\mathbf{V}/\partial\alpha_i\partial\alpha_j, \ldots. \tag{2·31}$$

Also
$$\mathbf{X}^2 = \Sigma\epsilon_i\epsilon_j\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j - \ldots. \tag{2·32}$$

Using the abbreviated notation

$$(i) = \operatorname{tr}\mathbf{V}^{-1}\mathbf{V}_i, \quad (i,j) = \operatorname{tr}\mathbf{V}^{-1}\mathbf{V}_{i,j}, \quad (i\,|\,j) = \operatorname{tr}\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j, \ldots, \tag{2·33}$$

we have
$$\operatorname{tr}\mathbf{X} = -\Sigma\epsilon_i(i) + \Sigma\epsilon_i\epsilon_j(i\,|\,j) - \ldots, \tag{2·34}$$

$$\operatorname{tr}\mathbf{X}^2 = \Sigma\epsilon_i\epsilon_j(i\,|\,j) - \ldots, \tag{2·35}$$

$$(\operatorname{tr}\mathbf{X})^2 = \Sigma\epsilon_i\epsilon_{j(i)}(j) - \ldots, \tag{2·36}$$

whence
$$J = [1 + \Sigma\epsilon_i\{-\tfrac{1}{2}(i)\Delta\} + \Sigma\epsilon_i\epsilon_j\{(i\,|\,j)(\tfrac{1}{2}\Delta + \tfrac{1}{4}\Delta^2) + \tfrac{1}{8}(i)(j)\Delta^2\} + \ldots]G_\rho(\xi). \tag{2·37}$$

Comparing (2·17) and (2·37) we obtain

$$\partial_i \Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha) \leqslant 2\xi] = -\tfrac{1}{2}(i)\Delta G_\rho(\xi) = \tfrac{1}{2}(i)Eg_\rho(\xi), \tag{2·38}$$

$$\begin{aligned}\partial_i\partial_j \Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha) \leqslant 2\xi] &= [(i\,|\,j)(\Delta + \tfrac{1}{2}\Delta^2) + \tfrac{1}{4}(i)(j)\Delta^2]G_\rho(\xi) \\ &= [-\tfrac{1}{2}(i\,|\,j)(E^2+E) - \tfrac{1}{4}(i)(j)(E^2-E)]g_\rho(\xi),\end{aligned} \tag{2·39}$$

where
$$g_\rho(\xi) = G'_\rho(\xi) = [\Gamma(\rho)]^{-1}\xi^{\rho-1}e^{-\xi}. \tag{2·40}$$

By a straightforward but tedious extension of the algebra we can obtain

$$\begin{aligned}\partial_i\partial_j\partial_k \Pr[\ldots] = [&(i\,|\,j\,|\,k)(E^3+E^2+E) + \tfrac{1}{2}((i\,|\,j)(k) + (i\,|\,k)(j) + (j\,|\,k)(i))(E^3-E) \\ &+ \tfrac{1}{8}(i)(j)(k)(E^3-2E^2+E)]g_\rho(\xi),\end{aligned} \tag{2·41}$$

$$\begin{aligned}\partial_i\partial_j\partial_k\partial_l \Pr[\ldots] = [&-((i\,|\,j\,|\,k\,|\,l) + (i\,|\,k\,|\,l\,|\,j) + (i\,|\,l\,|\,j\,|\,k))(E^4+E^3+E^2+E) \\ &- \tfrac{1}{2}((i\,|\,j\,|\,k)(l) + (i\,|\,j\,|\,l)(k) + (i\,|\,k\,|\,l)(j) + (j\,|\,k\,|\,l)(i))(E^4-E) \\ &- \tfrac{1}{4}((i\,|\,j)(k\,|\,l) + (i\,|\,k)(j\,|\,l) + (i\,|\,l)(j\,|\,k))(E^4+E^3-E^2-E) \\ &- \tfrac{1}{8}((i\,|\,j)(k)(l) + (i\,|\,k)(j)(l) + \ldots \text{ to 6 terms})(E^4-E^3-E^2+E) \\ &- \tfrac{1}{16}(i)(j)(k)(l)(E^4-3E^3+3E^2-E)]g_\rho(\xi).\end{aligned} \tag{2·42}$$

There is, however, one rather tiresome point that arises in this sort of work. If, for example, we have

$$\Sigma a_{ijk}\epsilon_i\epsilon_j\epsilon_k = \Sigma b_{ijk}\epsilon_i\epsilon_j\epsilon_k, \tag{2·43}$$

we are only entitled to infer that $a_{ijk} = b_{ijk}$ if both these sets of quantities are completely symmetrical in their suffixes. Now a trace such as $(i\,|\,j\,|\,k)$ is unaltered only by *cyclic permutations* and *reversals* of the order of its letters; thus $(i\,|\,j\,|\,k)$ is completely symmetrical in its letters, but $(i\,|\,j\,|\,k\,|\,l)$ and $(i\,|\,j)(k)$ are not, and must be replaced by

$$\tfrac{1}{3}((i\,|\,j\,|\,k\,|\,l) + (i\,|\,k\,|\,l\,|\,j) + (i\,|\,l\,|\,j\,|\,k)) \quad \text{and} \quad \tfrac{1}{2}((i\,|\,j)(k) + (i\,|\,k)(j) + (j\,|\,k)(i))$$

respectively. This has been done in (2·41) and (2·42).

Substituting (2·39) into (2·14) we obtain

$$h_1(\alpha) = \tfrac{1}{4}\Sigma\frac{\alpha_i^2}{\nu_i}\left[2\left(\frac{\xi^2}{\rho(\rho+1)} + \frac{\xi}{\rho}\right)(i\,|\,i) + \left(\frac{\xi^2}{\rho(\rho+1)} - \frac{\xi}{\rho}\right)(i)^2\right]. \tag{2·44}$$

Differentiating (2·44) twice with respect to α_i, and substituting the results, together with (2·38), (2·39), (2·41) and (2·42) into (2·15), we finally obtain, as the approximation of order ν^{-2},

$$\begin{aligned}
2h(a) = \chi^2 &+ \tfrac{1}{2}\Sigma\frac{a_i^2}{\nu_i}[2(\chi_4+\chi_2)(i\,|\,i)+(\chi_4-\chi_2)(i)^2] \\
&+\left\{\frac{1}{16}\left(1-\frac{r-2}{\chi^2}\right)\left[\Sigma\frac{a_i^2}{\nu_i}[2(\chi_4+\chi_2)(i\,|\,i)+(\chi_4-\chi_2)(i)^2]\right]^2\right. \\
&-\Sigma\frac{a_i^2}{\nu_i^2}[2(\chi_4+\chi_2)(i\,|\,i)+(\chi_4-\chi_2)(i)^2] \\
&+4\Sigma\frac{a_i^3}{\nu_i^2}[2(\chi_4+\chi_2)(i\,|\,i\,|\,i)+(\chi_4-\chi_2)(i\,|\,i)(i)] \\
&-\Sigma\frac{a_i^2a_j^2}{\nu_i\nu_j}[2(\chi_4+\chi_2)(2(i\,|\,i\,|\,j\,|\,j)+(i\,|\,j\,|\,i\,|\,j))+(\chi_4-\chi_2)(2(i\,|\,i\,|\,j)(j)+(i\,|\,j)^2)] \\
&+(\chi_2-1)\Sigma\frac{a_i^3}{\nu_i^2}[2(\chi_4+\chi_2)(i\,|\,i)(i)+(\chi_4-\chi_2)(i)^3] \\
&-(\chi_2-1)\Sigma\frac{a_i^2a_j^2}{\nu_i\nu_j}[2(\chi_4+\chi_2)(i\,|\,i\,|\,j)(j)+(\chi_4-\chi_2)(i\,|\,j)(i)(j)] \\
&-\tfrac{1}{8}\Sigma\frac{a_i^2a_j^2}{\nu_i\nu_j}[2(\chi_4-1)(i\,|\,i)+(\chi_4-2\chi_2+1)(i^2)][2(\chi_4+\chi_2)(j\,|\,j)+(\chi_4-\chi_2)(j)^2] \\
&-\tfrac{1}{3}\Sigma\frac{a_i^3}{\nu_i^2}[8(\chi_6+\chi_4+\chi_2)(i\,|\,i\,|\,i)+6(\chi_6-\chi_2)(i\,|\,i)(i)+(\chi_6-2\chi_4+\chi_2)(i)^3] \\
&+\tfrac{1}{16}\Sigma\frac{a_i^2a_j^2}{\nu_i\nu_j}[16(\chi_8+\chi_6+\chi_4+\chi_2)(2(i\,|\,i\,|\,j\,|\,j)+(i\,|\,j\,|\,i\,|\,j)) \\
&\qquad+32(\chi_8-\chi_2)(i\,|\,i\,|\,j)(j)+4(\chi_8+\chi_6-\chi_4-\chi_2)((i\,|\,i)(j\,|\,j)+2(i\,|\,j)^2) \\
&\qquad+4(\chi_8-\chi_6-\chi_4+\chi_2)((i\,|\,i)(j)^2+2(i\,|\,j)(i)(j)) \\
&\qquad\left.+(\chi_8-3\chi_6+3\chi_4-\chi_2)(i)^2(j)^2]\right\}+O(\nu^{-3}). \qquad (2\cdot45)
\end{aligned}$$

Here χ_{2s} denotes $\chi^{2s}/r(r+2)\ldots(r+2s-2)$, while of course (i) now denotes $\operatorname{tr}\mathbf{V}^{-1}(a)\,\mathbf{V}_i(a)$, and so on. The first line of this result contains the terms of order 1 and ν^{-1}, while the curled bracket contains the term of order ν^{-2}.

We now take up the *general case*. From (2·2),

$$\hat{\boldsymbol{\theta}}(\alpha+\epsilon) = \mathbf{C}^{-1}(\alpha+\epsilon)\,\mathbf{d}(\alpha+\epsilon) = \mathbf{C}^{-1}(\alpha+\epsilon)\,\mathbf{B}'(\boldsymbol{\alpha}+\boldsymbol{\epsilon})^{-1}\,\mathbf{x}. \qquad (2\cdot46)$$

Hence

$$\begin{aligned}
\operatorname{var}\hat{\boldsymbol{\theta}}(\alpha+\epsilon) &= \mathbf{C}^{-1}(\alpha+\epsilon)\,\mathbf{B}'(\boldsymbol{\alpha}+\boldsymbol{\epsilon})^{-1}\,\boldsymbol{\alpha}(\boldsymbol{\alpha}+\boldsymbol{\epsilon})^{-1}\,\mathbf{B}\mathbf{C}^{-1}(\alpha+\epsilon) \\
&= \mathbf{C}^{-1}(\alpha+\epsilon)\,\mathbf{C}((\alpha+\epsilon)^2/\alpha)\,\mathbf{C}^{-1}(\alpha+\epsilon), \qquad (2\cdot47)
\end{aligned}$$

and $\operatorname{var}\mathbf{t}(\alpha+\epsilon)$ is the matrix consisting or the leading r rows and columns of this, or say

$$\operatorname{var}\mathbf{t}(\alpha+\epsilon) = [\mathbf{C}^{-1}(\alpha+\epsilon)\,\mathbf{C}((\alpha+\epsilon)^2/\alpha)\,\mathbf{C}^{-1}(\alpha+\epsilon)]_{11}. \qquad (2\cdot48)$$

Thus, from (2·27),

$$\mathbf{X} = \mathbf{V}^{-1}(\alpha+\epsilon)\,[\boldsymbol{\Gamma}(\alpha+\epsilon)\,\boldsymbol{\Gamma}^{-1}((\alpha+\epsilon)^2/\alpha)\,\boldsymbol{\Gamma}(\alpha+\epsilon)]_{11}-\mathbf{I}, \qquad (2\cdot49)$$

where we have introduced a new symbol $\boldsymbol{\Gamma} = \mathbf{C}^{-1}$. Now

$$\begin{aligned}\boldsymbol{\Gamma}^{-1}((\alpha+\epsilon)^2/\alpha) &= \boldsymbol{\Gamma}^{-1}(\alpha+2\epsilon+\epsilon^2/\alpha)\\ &= \left[\boldsymbol{\Gamma}+\Sigma\left(2\epsilon_i+\frac{\epsilon_i^2}{\alpha_i}\right)\boldsymbol{\Gamma}_i+2\Sigma\epsilon_i\epsilon_j\boldsymbol{\Gamma}_{i,j}+\ldots\right]^{-1}\\ &= \left[\mathbf{I}+2\Sigma\epsilon_i\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_i+\Sigma\epsilon_i\epsilon_j\left(\frac{\delta_{ij}}{\alpha_i}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_i+2\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_{i,j}\right)+\ldots\right]^{-1}\boldsymbol{\Gamma}^{-1}\\ &= \left[\mathbf{I}-2\Sigma\epsilon_i\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_i+\Sigma\epsilon_i\epsilon_j\left(-\frac{\delta_{ij}}{\alpha_i}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_i-2\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_{i,j}+4\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_j\right)+\ldots\right]\boldsymbol{\Gamma}^{-1},\end{aligned} \tag{2·50}$$

where $\boldsymbol{\Gamma}$ denotes $\boldsymbol{\Gamma}(\alpha)$, $\boldsymbol{\Gamma}_i = \partial\boldsymbol{\Gamma}/\partial\alpha_i$, and so forth. Thus

$$\begin{aligned}\boldsymbol{\Gamma}(\alpha+\epsilon)\,\boldsymbol{\Gamma}^{-1}((\alpha+\epsilon)^2/\alpha)\,\boldsymbol{\Gamma}(\alpha+\epsilon) &= \boldsymbol{\Gamma}[\mathbf{I}+\Sigma\epsilon_i\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_i+\tfrac{1}{2}\Sigma\epsilon_i\epsilon_j\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_{i,j}+\ldots]\\ &\quad\times\left[\mathbf{I}-2\Sigma\epsilon_i\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_i+\Sigma\epsilon_i\epsilon_j\left(-\frac{\delta_{ij}}{\alpha_i}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_i-2\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_{i,j}+4\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_j\right)+\ldots\right]\\ &\quad\times[\mathbf{I}+\Sigma\epsilon_i\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_i+\tfrac{1}{2}\Sigma\epsilon_i\epsilon_j\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_{i,j}+\ldots]\\ &= \boldsymbol{\Gamma}+\Sigma\epsilon_i\epsilon_j\left(-\frac{\delta_{ij}}{\alpha_i}\boldsymbol{\Gamma}_i-\boldsymbol{\Gamma}_{i,j}+\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_j\right)+\ldots.\end{aligned} \tag{2·51}$$

Hence, since $\boldsymbol{\Gamma}_{11}$ (the leading r rows and columns of $\boldsymbol{\Gamma}$) is $\mathbf{V}$, (2·49) becomes

$$\begin{aligned}\mathbf{X} &= [\mathbf{I}+\Sigma\epsilon_i\mathbf{V}^{-1}\mathbf{V}_i+\tfrac{1}{2}\Sigma\epsilon_i\epsilon_j\mathbf{V}^{-1}\mathbf{V}_{i,j}+\ldots]^{-1}\mathbf{V}^{-1}\\ &\quad\times\left[\mathbf{V}+\Sigma\epsilon_i\epsilon_j\left(-\frac{\delta_{ij}}{\alpha_i}\mathbf{V}_i-\mathbf{V}_{i,j}+(\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_j)_{11}\right)+\ldots\right]-\mathbf{I}\\ &= -\Sigma\epsilon_i\mathbf{V}^{-1}\mathbf{V}_i+\Sigma\epsilon_i\epsilon_j\left(-\frac{\delta_{ij}}{\alpha_i}\mathbf{V}^{-1}\mathbf{V}_i-\tfrac{3}{2}\mathbf{V}^{-1}\mathbf{V}_{i,j}+\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j+\mathbf{V}^{-1}(\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_j)_{11}\right)+\ldots.\end{aligned} \tag{2·52}$$

The work of evaluating J and the derivatives now proceeds as before, and the final result is

$$\begin{aligned}2h(a) &= \chi^2-2\chi_2\Sigma\frac{a_i}{\nu_i}(i)+\Sigma\frac{a_i^2}{\nu_i}[-3\chi_2(i,i)+(\chi_4+\chi_2)(i\,|\,i)\\ &\quad+\tfrac{1}{2}(\chi_4-\chi_2)(i)^2+2\chi_2\,\mathrm{tr}\,\mathbf{V}^{-1}(\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_i)_{11}]+O(\nu^{-2}).\end{aligned} \tag{2·53}$$

Of course $\boldsymbol{\Gamma}$ now denotes $\boldsymbol{\Gamma}(a)$, and so forth.

In at least one application it is preferable to write the result (2·53) in terms of the matrices

$$\boldsymbol{\Lambda} = \mathbf{V}^{-1}, \quad \mathbf{C} = \boldsymbol{\Gamma}^{-1}, \tag{2·54}$$

and their derivatives. It is easy to prove that

$$(i) = -\,\mathrm{tr}\,\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}_i, \tag{2·55}$$

$$(i,i) = -\,\mathrm{tr}\,\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}_{i,i}+2\,\mathrm{tr}\,(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}_i)^2, \tag{2·56}$$

$$(i\,|\,i) = \mathrm{tr}\,(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}_i)^2, \tag{2·57}$$

$$\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_i = \mathbf{C}^{-1}\mathbf{C}_i\mathbf{C}^{-1}\mathbf{C}_i\mathbf{C}^{-1}. \tag{2·58}$$

Hence

$$\begin{aligned}2h(a) &= \chi^2+2\chi_2\Sigma\frac{a_i}{\nu_i}\mathrm{tr}\,\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}_i\\ &\quad+\Sigma\frac{a_i^2}{\nu_i}[3\chi_2\,\mathrm{tr}\,\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}_{i,i}+(\chi_4-5\chi_2)\,\mathrm{tr}\,(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}_i)^2+\tfrac{1}{2}(\chi_4-\chi_2)(\mathrm{tr}\,\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}_i)^2\\ &\quad+2\chi_2\,\mathrm{tr}\,\boldsymbol{\Lambda}(\mathbf{C}^{-1}\mathbf{C}_i\mathbf{C}^{-1}\mathbf{C}_i\mathbf{C}^{-1})_{11}]+O(\nu^{-2}).\end{aligned} \tag{2·59}$$

One of the formulae (2·53) or (2·59) replaces the first line of (2·45) in the *general case*. I have not worked out the term of order ν^{-2} in this case; it would obviously be extremely complicated, and is unlikely to be of practical use.

We may summarize the work as follows. To test the hypothesis $\theta_1 = \ldots = \theta_r = 0$, we first obtain the maximum-likelihood solution for $\boldsymbol{\theta}$ under the alternative hypothesis, namely, $\hat{\boldsymbol{\theta}} = \mathbf{C}^{-1}\mathbf{d}$, and at the same time its variance matrix $\mathbf{C}^{-1}$, *as though the estimates* a_i *were the true variances* α_i. Denote the vector formed by the first r elements of this solution by $\mathbf{t}$, and the corresponding part of the variance matrix by $\mathbf{V}$. If the a_i are based on sufficiently large numbers ν_i, of degree of freedom, then $\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}$ may be compared directly with the appropriate significance point of χ^2 on r degrees of freedom. If not, a more accurate test will be to compare it with $2h(a)$ given by (2·45), in the *simple case* and by (2·53) or (2·59) in the *general case*.

A slight generalization may be made in our initial assumptions, which may be useful in some applications. Instead of taking the variance of each x_i as α_i, we may assume that

$$\operatorname{var} x_i = \sum_{j=1}^{m} \beta_{ij}\alpha_j, \tag{2·60}$$

where $\boldsymbol{\beta} = (\beta_{ij})$, like $\mathbf{B}$ of (2·1), is a matrix of known constants. The α_j are estimated by quantities a_j as before, and the only alteration necessary in our work is that (2·3) must be replaced by

$$\mathbf{C} = \mathbf{B}'(\operatorname{var}\mathbf{x})^{-1}\mathbf{B}, \quad \mathbf{d} = \mathbf{B}'(\operatorname{var}\mathbf{x})^{-1}\mathbf{x}, \tag{2·61}$$

where $\operatorname{var}\mathbf{x}$ is a diagonal matrix whose diagonal elements are given by the previous equation. Of course $\mathbf{C}$ and $\mathbf{d}$ are still functions of the α_i, so that we are still entitled to the notation $\mathbf{t}(\alpha)$, etc.

The commonest use of this generalization may well be in problems where some of the x_i may be assumed to have the same variance, or known multiples of a common variance; such a situation could arise in a two-factor lay-out (see § 4) in which the error variance only depended on the level of one factor, being independent of the other.

An interesting special case of (2·60) occurs when there is only one α; we may then write $\operatorname{var} x_i = \beta_i\alpha$, say, and the hypothesis reverts to the classical form in which there is only one variance, α, to be estimated, the quantities β_i^{-1} being known 'weights'. The statistic $\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}$ reduces to r times the quantity taken to be distributed as F with r and ν degrees of freedom in the classical case, and

$$2h(a) = \chi^2\left[1 + \frac{\chi^2 - (r-2)}{2\nu} + \frac{4\chi^4 - 11(r-2)\chi^2 + (r-2)(7r-10)}{24\nu^2} + O(\nu^{-3})\right]. \tag{2·62}$$

This formula, then, is an expansion of r times the significance point of an F distribution with r and ν degrees of freedom, in terms of the corresponding significance point of a χ^2 distribution with r degrees of freedom. The first two terms were given by Welch (1951).

3. Simple examples of the univariate linear hypothesis

The simplest example one can think of is that in which the structural relations (2·1) are simply

$$\mathscr{E}(x_i) = \theta_i \quad (i = 1, \ldots, k), \tag{3·1}$$

and the null hypothesis to be tested is that

$$\theta_1 = \ldots = \theta_k = 0. \tag{3·2}$$

This arises if the x_i are the means of k samples, and the null hypothesis states that the corresponding population means have specified values. No generality is lost by taking these values to be zero (see footnote on p. 19).

In this case the maximum-likelihood solution for the θ_i under the alternative hypothesis is

$$t_i = x_i \quad (i = 1, \ldots, k), \tag{3·3}$$

whence the variance matrix of $\mathbf{t}$ is

$$\mathbf{V} = \operatorname{diag}(a_1, \ldots, a_k). \tag{3·4}$$

Since $\mathbf{t}$ does not depend on the variances, we have an example of the *simple case*, to which (2·45) applies. Writing

$$w_i = 1/a_i \quad (i = 1, \ldots, k) \tag{3·5}$$

we have

$$\mathbf{V}^{-1} = \operatorname{diag}(w_1, \ldots, w_k), \tag{3·6}$$

$$\mathbf{V}_i = \operatorname{diag}(0, \ldots, 0, 1, 0, \ldots, 0), \tag{3·7}$$

having unity in the ith place only, so that

$$(i \mid i) = (i)^2 = w_i^2, \tag{3·8}$$

$$(i \mid i \mid i) = (i \mid i)\,(i) = (i)^3 = w_i^3, \tag{3·9}$$

$$(i \mid i \mid j \mid j) = (i \mid j \mid i \mid j) = (i \mid i \mid j)\,(j) = (i \mid j)^2 = (i \mid j)\,(i)\,(j) = \delta_{ij} w_i^4, \tag{3·10}$$

$$(i \mid i)\,(j \mid j) = (i \mid i)\,(j)^2 = (i)^2\,(j)^2 = w_i^2 w_j^2. \tag{3·11}$$

Thus the two quantities to be compared for test purposes are

$$\mathbf{t}'\mathbf{V}^{-1}\mathbf{t} = \Sigma w_i x_i^2 \tag{3·12}$$

and

$$\begin{aligned} 2h(a) = {} & \chi^2 + \tfrac{1}{2}(3\chi_4 + \chi_2)\,\Sigma\nu_i^{-1} \\ & + \left\{ 2(3\chi_8 - \chi_6)\,\Sigma\nu_i^{-2} + \frac{1}{16}\left[\left(1 - \frac{k-2}{\chi^2}\right)(3\chi_4 + \chi_2)^2 \right.\right. \\ & \left.\left. - 2(3\chi_4 - 2\chi_2 - 1)\,(3\chi_4 + \chi_2) + (9\chi_8 - 3\chi_6 - 5\chi_4 - \chi_2) \right] (\Sigma\nu_i^{-1})^2 \right\} + O(\nu^{-3}). \end{aligned} \tag{3·13}$$

χ^2 is based on k degrees of freedom and χ_{2s} denotes $\chi^{2s}/k(k+2)\ldots(k+2s-2)$.

It will be noticed that in this example $2h(a)$ does not actually depend on the a_i; this is because $\Sigma w_i x_i^2$ itself has a distribution which is independent of the unknown variances; it is, in fact, the sum of k quantities distributed independently in F distributions with $(1, \nu_1), \ldots, (1, \nu_k)$ degrees of freedom.

As a second example, suppose that we have

$$\mathscr{E}(x_i) = \mu_i \quad (i = 1, \ldots, k), \tag{3·14}$$

and that the null hypothesis is

$$\mu_1 = \ldots = \mu_k, \tag{3·15}$$

that is, that the means are all equal, their common value not being specified. This hypothesis, which only imposes $(k-1)$ constraints, is not in the canonical form, but is easily put into this form by means of the transformation

$$\mu_i = \theta_i + \theta_k \quad (i = 1, \ldots, k-1), \tag{3·16}$$

$$\mu_k = \theta_k, \tag{3·17}$$

when it becomes

$$\theta_1 = \ldots = \theta_{k-1} = 0. \tag{3·18}$$

The maximum-likelihood solution for $\boldsymbol{\theta}$ under the alternative hypothesis gives us

$$\mathbf{t} = \{x_1 - x_k, \ldots, x_{k-1} - x_k\}, \tag{3·19}$$

whence
$$\mathbf{V} = \begin{bmatrix} a_1+a_k & a_k & \dots & a_k \\ a_k & a_2+a_k & \dots & a_k \\ \dots & \dots & \dots & \dots \\ a_k & a_k & \dots & a_{k-1}+a_k \end{bmatrix}, \tag{3·20}$$

$$\mathbf{V}^{-1} = \begin{bmatrix} w_1-w_1^2/w & -w_1w_2/w & \dots & -w_1w_{k-1}/w \\ -w_2w_1/w & w_2-w_2^2/w & \dots & -w_2w_{k-1}/w \\ \dots & \dots & \dots & \dots \\ -w_{k-1}w_1/w & -w_{k-1}w_2/w & \dots & w_{k-1}-w_{k-1}^2/w \end{bmatrix}, \tag{3·21}$$

where
$$w_i = 1/a_i, \quad w = \sum_1^k w_i. \tag{3·22}$$

Thus
$$\mathbf{V}_i = \operatorname{diag}(0, \dots, 0, 1, 0, \dots, 0) \quad (i = 1, \dots, k-1), \tag{3·23}$$

$$\mathbf{V}_k = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & 1 \end{bmatrix}, \tag{3·24}$$

and
$$(i \mid i) = (i^2) = (w_i - w_i^2/w)^2, \tag{3·25}$$
$$(i \mid i \mid i) = (i \mid i)\,(i) = (i)^3 = (w_i - w_i^2/w)^3, \tag{3·26}$$
$$(i \mid i \mid j \mid j) = (i \mid i \mid j)\,(j) = (i \mid j)\,(i)\,(j) = \left(w_i - \frac{w_i^2}{w}\right)\left(w_j - \frac{w_j^2}{w}\right)\left(\delta_{ij}w_i - \frac{w_iw_j}{w}\right)^2, \tag{3·27}$$
$$(i \mid j \mid i \mid j) = \left(\delta_{ij}w_i - \frac{w_iw_j}{w}\right)^4, \tag{3·28}$$
$$(i \mid i)\,(j \mid j) = (i \mid i)\,(j)^2 = (i)^2\,(j)^2 = \left(w_i - \frac{w_i^2}{w}\right)^2\left(w_j - \frac{w_j^2}{w}\right)^2. \tag{3·29}$$

(These latter results hold for $i, j = 1, \dots, k$, but they are most easily proved when $i, j \neq k$, the results when i or $j = k$ following on consideration of the fact that all the traces involved are left invariant by non-singular linear transformations of the t_i; thus in spite of the special part played by x_k in (3·19), it could be replaced by any other x without altering the results (3·25)–(3·29), so that the latter cannot be restricted to the case $i, j \neq k$.)

Since $\mathbf{t}$ does not depend on the variances, we again use (2·45), which gives

$$\begin{aligned} 2h(a) = {} & \chi^2 + \tfrac{1}{2}(3\chi_4+\chi_2)\,\Sigma\frac{1}{\nu_i}\left(1-\frac{w_i}{w}\right)^2 \\ & + \left\{\frac{1}{16}\left(1-\frac{k-3}{\chi^2}\right)\left[(3\chi_4+\chi_2)\,\Sigma\frac{1}{\nu_i}\left(1-\frac{w_i}{w}\right)^2\right]^2\right. \\ & + \tfrac{1}{8}(3\chi_4+\chi_2)\,[8(2R_{21}-5R_{22}+4R_{23}-2R_{11}^2+4R_{11}R_{12}-3R_{12}^2) \\ & \quad + 8(\chi_2-1)\,(R_{21}-2R_{22}+R_{23}-R_{11}^2+2R_{11}R_{12}-R_{12}^2) \\ & \quad - (3\chi_4-2\chi_2-1)\,(R_{10}^2-4R_{10}R_{11}+2R_{10}R_{12}+4R_{11}^2-4R_{11}R_{12}+R_{12}^2)] \\ & - (5\chi_6+2\chi_4+\chi_2)\,(R_{20}-3R_{21}+3R_{22}-R_{23}) \\ & + \tfrac{1}{2}(9\chi_8+3\chi_6+3\chi_4+\chi_2)\,(R_{20}-4R_{21}+5R_{22}-2R_{23}+R_{11}^2-2R_{11}R_{12}+R_{12}^2) \\ & + \tfrac{1}{2}(3\chi_8+3\chi_6+\chi_4+\chi_2)\,(R_{20}-4R_{21}+6R_{22}-4R_{23}+R_{12}^2) \\ & \left. + \tfrac{1}{16}(9\chi_8-3\chi_6-5\chi_4-\chi_2)\,(R_{10}^2-4R_{10}R_{11}+2R_{10}R_{12}+4R_{11}^2-4R_{11}R_{12}+R_{12}^2)\right\} + O(\nu^{-3}), \end{aligned} \tag{3·30}$$

where
$$R_{st} = \Sigma \frac{1}{v_i^s}\left(\frac{w_i}{w}\right)^t. \tag{3·31}$$

χ^2 is based on $(k-1)$ degrees of freedom, and χ_{2s} denotes

$$\chi^{2s}/(k-1)(k+1)\dots(k+2s-3).$$

It is well known from large sample theory, although it may be proved directly from (3·19) and (3·21), that
$$\mathbf{t'V^{-1}t} = \Sigma w_i x_i^2 - (\Sigma w_i x_i)^2/w \tag{3·32}$$
in this problem. The test carried out by comparing (3·32) with (3·20) is the same as that obtained in the previous paper (although the formula for $2h(a)$ is arranged rather more compactly here).

4. Application to a two-factor experiment with replication

We suppose that we have a factorial experiment with two factors A and B, A being applied to p levels and B at q levels, and that at the ith level of A and the jth level of B we have $\nu_{ij}+1$ replicate determinations. We denote the individual results by x_{ijl} $(i = 1,\dots,p;\ j = 1,\dots,q;\ l = 1,\dots,\nu_{ij}+1)$ and assume that

$$\mathscr{E}(x_{ijl}) = \mu+\alpha_i+\beta_j+\gamma_{ij} \tag{4·1}$$

with
$$\Sigma\alpha_i = \Sigma\beta_j = \Sigma_i\gamma_{ij} = \Sigma_j\gamma_{ij} = 0. \tag{4·2}$$

We also assume that the deviations from these expectations are independently and normally distributed about zero, with variances which may differ from treatment combination to treatment combination, but which are constant for all $\nu_{ij}+1$ observations with treatment combination (i, j). Let x_{ij} denote the mean of these $\nu_{ij}+1$ observations and let a_{ij} denote the estimate of its variance, based on ν_{ij} degrees of freedom, derived from the 'within treatments' sums of squares. That is,

$$x_{ij} = \frac{\Sigma_l x_{ijl}}{\nu_{ij}+1}, \quad a_{ij} = \frac{\Sigma_l(x_{ijl}-x_{ij})^2}{\nu_{ij}(\nu_{ij}+1)} \quad (i = 1,\dots,p;\ j = 1,\dots,q). \tag{4·3}$$

An overall test for the significance of all the treatment effects (main effects and interactions) is afforded by treating the x_{ij} as $k = pq$ sample means, as considered in the previous section. We now consider the more commonly required tests for various groups of constants.

(a) *Test for interaction*

The null hypothesis is that all the γ_{ij} are zero, the α_i and β_j remaining unspecified; the alternative hypothesis is that none of the parameters are specified (always subject, of course, to the conditions (4·2)).

In order to express this hypothesis in canonical form we write, in place of (4·1) and (4·2),

$$\mathscr{E}(x_{ij}) = \theta_{ij}+\phi_i+\psi_j+\omega \quad (i = 1,\dots,p-1;\ j = 1,\dots,q-1), \tag{4·4}$$

$$\mathscr{E}(x_{iq}) = \phi_i + \omega \quad (i = 1,\dots,p-1), \tag{4·5}$$

$$\mathscr{E}(x_{pj}) = \psi_j+\omega \quad (j = 1,\dots,q-1), \tag{4·6}$$

$$\mathscr{E}(x_{pq}) = \omega, \tag{4·7}$$

there being no restrictions corresponding to (4·2). The null hypothesis is now that all the θ_{ij} vanish.

The maximum-likelihood estimates of the θ_{ij} are

$$t_{ij} = x_{ij} - x_{iq} - x_{pj} + x_{pq} \quad (i = 1, \ldots, p-1;\ j = 1, \ldots, q-1), \tag{4·8}$$

and we have an example of the *simple case*. Hence the variance matrix of the vector $\mathbf{t}$ is

$$\mathbf{V} = \begin{bmatrix} a_{11}+a_{1q}+a_{p1}+a_{pq} & a_{1q}+a_{pq} & a_{p1}+a_{pq} & a_{pq} \\ a_{1q}+a_{pq} & a_{12}+a_{1q}+a_{p2}+a_{pq} & a_{pq} & a_{p2}+a_{pq} \\ a_{p1}+a_{pq} & a_{pq} & a_{21}+a_{2q}+a_{p1}+a_{pq} & a_{2q}+a_{pq} \\ a_{pq} & a_{p2}+a_{pq} & a_{2q}+a_{pq} & a_{22}+a_{2q}+a_{p2}+a_{pq} \end{bmatrix}, \tag{4·9}$$

where, for simplicity, we have taken the case $p = q = 3$; the elements of $\mathbf{t}$ are arranged in dictionary order, thus: $\{t_{11}, t_{12}, t_{21}, t_{22}\}$. This is easily differentiated to give the first-order derivatives $\mathbf{V}_{ij} = \partial \mathbf{V}/\partial a_{ij}$:

$$\mathbf{V}_{11} = \operatorname{diag}(1, 0, 0, 0), \quad \mathbf{V}_{12} = \operatorname{diag}(0, 1, 0, 0),$$
$$\mathbf{V}_{21} = \operatorname{diag}(0, 0, 1, 0), \quad \mathbf{V}_{22} = \operatorname{diag}(0, 0, 0, 1), \tag{4·10}$$

$$\mathbf{V}_{1q} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{V}_{2q} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \tag{4·11}$$

$$\mathbf{V}_{p1} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{V}_{p2} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \tag{4·12}$$

$$\mathbf{V}_{pq} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}. \tag{4·13}$$

The author has not succeeded in finding a simple form for $\mathbf{V}^{-1}$, corresponding to (3·21) of the last section, but it can of course be calculated numerically in any given case (see also below), and we shall denote its (ij, kl)th element by $v^{ij,kl}$. Then

$$(ij \mid ij) = (ij)^2 = (v^{ij,ij})^2 \quad (i = 1, \ldots, p-1;\ j = 1, \ldots, q-1), \tag{4·14}$$

$$(iq \mid iq) = (iq)^2 = (v^{i\cdot,i\cdot})^2 \quad (i = 1, \ldots, p-1), \tag{4·15}$$

$$(pj \mid pj) = (pj)^2 = (v^{\cdot j,\cdot j})^2 \quad (j = 1, \ldots, q-1), \tag{4·16}$$

$$(pq \mid pq) = (pq)^2 = (v^{\cdots\cdot})^2, \tag{4·17}$$

where $(ij \mid ij)$ denotes $\operatorname{tr}(\mathbf{V}^{-1}\mathbf{V}_{ij})^2$, etc., and the dots denote summation over the indices omitted, from 1 up to $(p-1)$ or from 1 up to $(q-1)$ as the case may be. If we make the convention that an index p or q, as the case may be, is to have this same meaning of summation, then (4·14)–(4·17) can be written as one formula:

$$(ij \mid ij) = (ij)^2 = (v^{ij,ij})^2 \quad (i = 1, \ldots, p;\ j = 1, \ldots, q). \tag{4·18}$$

Substituting into (2·45),

$$2h(a) = \chi^2 + \tfrac{1}{2}(3\chi_4 + \chi_2) \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(a_{ij} v^{ij,ij})^2}{\nu_{ij}} + O(\nu^{-2}), \tag{4·19}$$

the terms with $i = p$ or $j = q$ being interpreted according to the above convention. χ^2 is based on $r = (p-1)(q-1)$ degrees of freedom, and χ_{2s} denotes $\chi^{2s}/(r+2)\dots(r+2s-2)$. (The author has evaluated the next term in the expansion, but it is not thought to be of sufficient interest to give the rather complicated formula here.)

The quantity $\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}$ can either be calculated directly, once $\mathbf{V}^{-1}$ has been calculated, or as the difference between the (weighted) sum of squares due to fitting all the constants and that due to fitting the main effects only. If we denote the partitioning of the various matrices involved according to their first r $(=(p-1)(q-1))$ and last $k-r$ $(=p+q-1)$ rows and/or columns by the numbers 1 and 2, and in particular if

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}, \quad \mathbf{C}^{-1} = \begin{bmatrix} \mathbf{C}^{11} & \mathbf{C}^{12} \\ \mathbf{C}^{21} & \mathbf{C}^{22} \end{bmatrix}, \tag{4·20}$$

then this last method used the identity

$$\begin{aligned} \mathbf{t}'\mathbf{V}^{-1}\mathbf{t} \equiv \hat{\boldsymbol{\theta}}_1'(\mathbf{C}^{11})^{-1}\hat{\boldsymbol{\theta}}_1 &= \hat{\boldsymbol{\theta}}'\mathbf{C}\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_2^{*\prime}\mathbf{C}_{22}^{-1}\hat{\boldsymbol{\theta}}_2^* \\ &= \hat{\boldsymbol{\theta}}'\mathbf{d} - \hat{\boldsymbol{\theta}}_2^{*\prime}\mathbf{d}_2, \end{aligned} \tag{4·21}$$

where $\hat{\boldsymbol{\theta}}_2^*$ is the maximum-likelihood solution for $\boldsymbol{\theta}_2$ under the null hypothesis $\boldsymbol{\theta}_1 = 0$. If we adopt this well-known method of calculating $\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}$ (and it is the more economical one unless p and q are quite small), then we can evaluate $\mathbf{C}_{22}^{-1}$ at the same time, and obtain $\mathbf{V}^{-1}$ from the identity

$$\mathbf{V}^{-1} = (\mathbf{C}^{11})^{-1} = \mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}, \tag{4·22}$$

rather than by the direct inversion of $\mathbf{V}$.

(*b*) *Test for a main effect*

We now assume that the interaction constants are zero, so that (4·4)–(4·7) become, after a trivial change of notation,

$$\mathscr{E}(x_{ij}) = \phi_i + \omega_j \quad (i = 1, \dots, p-1;\ j = 1, \dots, q), \tag{4·23}$$

$$\mathscr{E}(x_{pj}) = \omega_j \quad (j = 1, \dots, q). \tag{4·24}$$

The null hypothesis is that the ϕ_i vanish. The matrices $\mathbf{C}$, $\boldsymbol{\theta}$ and $\mathbf{d}$ are (for the case $p = q = 3$)

$$\mathbf{C} = \left[\begin{array}{cc:ccc} w_{1.} & . & w_{11} & w_{12} & w_{1q} \\ . & w_{2.} & w_{21} & w_{22} & w_{2q} \\ \hdashline w_{11} & w_{21} & w_{.1} & . & . \\ w_{12} & w_{22} & . & w_{.2} & . \\ w_{1q} & w_{2q} & . & . & w_{.q} \end{array}\right], \quad \boldsymbol{\theta} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \hdashline \omega_1 \\ \omega_2 \\ \omega_q \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} w_{1.}\bar{x}_{1.} \\ w_{2.}\bar{x}_{2.} \\ \hdashline w_{.1}\bar{x}_{.1} \\ w_{.2}\bar{x}_{.2} \\ w_{.q}\bar{x}_{.q} \end{bmatrix}, \tag{4·25}$$

where

$$w_{ij} = 1/a_{ij}, \quad w_{i.} = \sum_{j=1}^{q} w_{ij}, \quad w_{.j} = \sum_{i=1}^{p} w_{ij}, \tag{4·26}$$

and

$$w_{i.}\bar{x}_{i.} = \sum_{j=1}^{q} w_{ij}x_{ij}, \quad w_{.j}\bar{x}_{.j} = \sum_{i=1}^{p} w_{ij}x_{ij}. \tag{4·27}$$

Since the solution $\mathbf{t}$ is obviously a function of the a_{ij}, this linear hypothesis is an example of the *general case*, and we must not be surprised if the application of the formula (2·59) gives a rather complicated test, even as far as the term of order ν^{-1}.

From (4·25),

$$\mathbf{\Lambda} \equiv \mathbf{V}^{-1} = \begin{bmatrix} w_{1.} & & \cdot \\ & \ddots & \\ \cdot & & w_{q.} \end{bmatrix} - \begin{bmatrix} w_{11} & w_{12} & w_{1q} \\ w_{21} & w_{22} & w_{2q} \end{bmatrix} \begin{bmatrix} w_{.1} & & \cdot \\ & w_{.2} & \\ \cdot & & w_{.q} \end{bmatrix}^{-1} \begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \\ w_{1q} & w_{2q} \end{bmatrix}$$

$$= \begin{bmatrix} w_{1.} - \Sigma w_{1l}^2/w_{.l} & -\Sigma w_{1l}w_{2l}/w_{.l} \\ -\Sigma w_{2l}w_{1l}/w_{.l} & w_{2.} - \Sigma w_{2l}^2/w_{.l} \end{bmatrix}, \tag{4·28}$$

or

$$\lambda_{rs} \equiv v^{rs} = \delta_{rs}w_{r.} - \sum_{l=1}^{q} w_{rl}w_{sl}/w_{.l}. \tag{4·29}$$

It readily follows that the typical elements of the first and second derivatives, $\mathbf{\Lambda}_{ij}$ and $\mathbf{\Lambda}_{ij,ij}$, are

$$\frac{\partial \lambda_{rs}}{\partial a_{ij}} = -w_{ij}^2 \frac{\partial \lambda_{rs}}{\partial w_{ij}} = -w_{ij}^2 \left(\delta_{ir} - \frac{w_{rj}}{w_{.j}}\right)\left(\delta_{is} - \frac{w_{sj}}{w_{.j}}\right), \tag{4·30}$$

$$\frac{\partial^2 \lambda_{rs}}{\partial a_{ij}^2} = 2w_{ij}^3 \left(1 - \frac{w_{ij}}{w_{.j}}\right)\left(\delta_{ir} - \frac{w_{rj}}{w_{.j}}\right)\left(\delta_{is} - \frac{w_{sj}}{w_{.j}}\right); \tag{4·31}$$

these hold for $i = 1, \ldots, p; j = 1, \ldots, q; r, s = 1, \ldots, p-1$. No explicit formula is known for $\mathbf{\Lambda}^{-1} = \mathbf{V}$, but it can of course be evaluated numerically. We now find

$$\operatorname{tr} \mathbf{\Lambda}^{-1}\mathbf{\Lambda}_{ij} = -w_{ij}u_{ij}, \tag{4·32}$$

$$\operatorname{tr} \mathbf{\Lambda}^{-1}\mathbf{\Lambda}_{ij,ij} = 2w_{ij}^2 \left(1 - \frac{w_{ij}}{w_{.j}}\right) u_{ij}, \tag{4·33}$$

$$\operatorname{tr} (\mathbf{\Lambda}^{-1}\mathbf{\Lambda}_{ij})^2 = (\operatorname{tr} \mathbf{\Lambda}^{-1}\mathbf{\Lambda}_{ij})^2 = w_{ij}^2 u_{ij}^2, \tag{4·34}$$

where

$$u_{ij} = w_{ij} \sum_{r,s=1}^{p-1} v_{rs} \left(\delta_{ir} - \frac{w_{rj}}{w_{.j}}\right)\left(\delta_{is} - \frac{w_{sj}}{w_{.j}}\right). \tag{4·35}$$

Finally, we have to evaluate $\operatorname{tr} \mathbf{\Lambda}(\mathbf{C}^{-1}\mathbf{C}_{ij}\mathbf{C}^{-1}\mathbf{C}_{ij}\mathbf{C}^{-1})_{11}$. Again, no explicit formula is known for $\mathbf{C}^{-1}$, and we shall denote it by

$$\mathbf{C}^{-1} = \left[\begin{array}{cc|ccc} c^{11} & c^{12} & c^{11}_{\cdot} & c^{12}_{\cdot} & c^{1q}_{\cdot} \\ c^{21} & c^{22} & c^{21}_{\cdot} & c^{22}_{\cdot} & c^{2q}_{\cdot} \\ \hline c^{11}_{\cdot} & c^{12}_{\cdot} & c^{11}_{\cdot\cdot} & c^{12}_{\cdot\cdot} & c^{1q}_{\cdot\cdot} \\ c^{21}_{\cdot} & c^{22}_{\cdot} & c^{21}_{\cdot\cdot} & c^{22}_{\cdot\cdot} & c^{2q}_{\cdot\cdot} \\ c^{q1}_{\cdot} & c^{q2}_{\cdot} & c^{q1}_{\cdot\cdot} & c^{q2}_{\cdot\cdot} & c^{qq}_{\cdot\cdot} \end{array}\right], \tag{4·36}$$

the asterisks meaning that the corresponding indices refer to the rows or columns belonging to the parameters ω_j. Some calculation now gives (for the case $p = q = 3$)

$$(\mathbf{C}^{-1}\mathbf{C}_{ij}\mathbf{C}^{-1}\mathbf{C}_{ij}\mathbf{C}^{-1})_{11} = w_{ij}^4 (c^{ii} + 2c^{ij}_{\cdot} + c^{jj}_{\cdot\cdot}) \begin{bmatrix} c^{1i} + c^{1j}_{\cdot} \\ c^{2i} + c^{2j}_{\cdot} \end{bmatrix} [c^{1i} + c^{1j}_{\cdot} \quad c^{2i} + c^{2j}_{\cdot}]$$

$$(i = 1, \ldots, p-1; j = 1, \ldots, q), \tag{4·37}$$

$$(\mathbf{C}^{-1}\mathbf{C}_{pj}\mathbf{C}^{-1}\mathbf{C}_{pj}\mathbf{C}^{-1})_{11} = w_{pj}^4 c^{jj}_{\cdot\cdot} \begin{bmatrix} c^{1j}_{\cdot} \\ c^{2j}_{\cdot} \end{bmatrix} [c^{1j}_{\cdot} \quad c^{2j}_{\cdot}] \quad (j = 1, \ldots, q). \tag{4·38}$$

If we agree to interpret the c symbols not occurring in (4·36) as zeros, (4·37) is also true for $i = p$. Combining this result with (4·29) we have

$$\operatorname{tr}\boldsymbol{\Lambda}(\mathbf{C}^{-1}\mathbf{C}_{ij}\mathbf{C}^{-1}\mathbf{C}_{ij}\mathbf{C}^{-1})_{11} = w_{ij}^4(c^{1i}+2c_{\cdot}^{1i}+c_{\cdot\cdot}^{1i})\,[c^{1i}+c_{\cdot}^{1i}\quad c^{2i}+c_{\cdot}^{2i}]\,\boldsymbol{\Lambda}\begin{bmatrix}c^{1i}+c_{\cdot}^{1i}\\ c^{2i}+c_{\cdot}^{2i}\end{bmatrix}$$

$$= w_{ij}^2 U_{ij} \quad (i = 1, \ldots, p;\ j = 1, \ldots, q) \tag{4·39}$$

where
$$U_{ij} = w_{ij}^2(c^{ii}+2c_{\cdot}^{ii}+c_{\cdot\cdot}^{ii})\sum_{r,s=1}^{p-1}\left(\delta_{rs}w_{r\cdot}-\sum_{l=1}^{q}\frac{w_{rl}w_{sl}}{w_{\cdot l}}\right)(c^{ri}+c_{\cdot}^{ri})\,(c^{si}+c_{\cdot}^{si}), \tag{4·40}$$

c's with no other meaning being interpreted as zero, as before.

Thus substituting (4·32), (4·33), (4·34), and (4·39) into (2·59) we obtain

$$2h(a) = \chi^2 + \Sigma\frac{1}{\nu_{ij}}\left[2\chi_2\left(2-3\frac{w_{ij}}{w_{\cdot j}}\right)u_{ij}+\tfrac{1}{2}(3\chi_4-11\chi_2)\,u_{ij}^2+2\chi_2 U_{ij}\right]+O(\nu^{-2}), \tag{4·41}$$

where χ^2 is based on $(p-1)$ degrees of freedom, χ_{2s} denotes $\chi^{2s}/(p-1)\,(p+1)\ldots(p+2s-3)$, and u_{ij}, U_{ij} are given by (4·35) and (4·40) respectively. (Since $\mathbf{V} \equiv \mathbf{C}^{11}$, the former equation may be written in terms of the c^{rs}, like the latter, if desired.)

Again there are two ways of calculating $\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}$: either directly, or as a difference between two sums of squares. In either case the greater part of the labour lies in computing the solution $\hat{\boldsymbol{\theta}}$, or at any rate the part $\hat{\boldsymbol{\theta}}_1 \equiv \mathbf{t}$. $\mathbf{V}^{-1}$ is readily found from (4·28)–(4·29), and $\mathbf{V}$ by inversion.

It should be pointed out that the occasions on which the results of this section may be used as they stand are likely to be few and far between. Second-order terms are either unknown or unquoted, and would in any case take a considerable time to compute. Consequently with only two or three replications $2h(a)$ will not be well determined. The eventual practical solution to this difficulty may lie in assuming that the variances are given by some such law as $\alpha_{ij} = \alpha_i' + \alpha_j''$ or $\alpha_{ij} = \alpha_i'\alpha_j''$, and then estimating the constants α_i', α_j'', which are fewer in number than the α_{ij}. Put in another way, we could use our present estimates a_{ij} to provide 'smoothed' estimates a_{ij}^* obeying such a law, thereby gaining (in an intuitive sense) degrees of freedom in our smoothed estimates. We would pay for this increase in that the distributions of the a_{ij}^* would be complicated and non-independent. This situation may not be quite as hopeless as it appears at first sight, for it may be shown that (reverting to the notation of § 2) whatever the joint distribution of the a_i is, provided that

$$\mathscr{E}(a_i) = \alpha_i + O(\nu^{-1}), \tag{4·42}$$

$$\operatorname{cov}(a_i, a_j) = O(\nu^{-1}), \tag{4·43}$$

and the higher cumulants are $o(\nu^{-1})$, we have

$$\Theta = 1 + \tfrac{1}{2}\Sigma\partial_i\partial_j \operatorname{cov}(a_i, a_j) + o(\nu^{-1}), \tag{4·44}$$

which reduces to (2·12) when the a_i have independent χ^2-type distributions. Thus it may be practicable to set up a first-order theory for this and other more complicated situations.

5. A MULTIVARIATE LINEAR HYPOTHESIS

Suppose that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are n p-variate vectors whose expected values are known linear functions of k unknown parameters $\theta_1, \ldots, \theta_k$, and that they are independently distributed around these expected values in p-variate normal distributions whose variance matrices are $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$. In matrix notation we again have

$$\mathscr{E}(\mathbf{x}) = \mathbf{B}\boldsymbol{\theta}, \tag{5·1}$$

but $\mathbf{x}$ now denotes the column vector $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, and $\mathbf{B}$, the matrix of known constants, is now of size $np \times k$. Let $\mathbf{A}_1, \ldots, \mathbf{A}_n$ be unbiased estimates of $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n$ distributed independently of the $\mathbf{x}_i$ and of each other in Wishart forms with $\nu_1, \ldots, \nu_n$ degrees of freedom.

The null hypothesis to be tested is, as before, that $\theta_1 = \ldots = \theta_r = 0$, where r is some number between 1 and k. If we knew the $\boldsymbol{\alpha}_i$ the appropriate test would again be based on the equation

$$\Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha) \leqslant 2\xi] = G_\rho(\xi), \tag{5·2}$$

the symbols having the same meanings as in § 2, except that $\boldsymbol{\alpha}$ in (2·3) now denotes $\mathbf{diag}\,(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n)$. As we do not know the $\boldsymbol{\alpha}_i$, we attempt to find a function $h(a)$ of the elements a_{irs} of the matrices $\mathbf{A}_i$, which is such that

$$\Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(a) \leqslant 2h(a)] = G_\rho(\xi). \tag{5·3}$$

Now
$$\Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(a) \leqslant 2h(a)] = \int \Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(a) \leqslant 2h(a) \mid da] \Pr[da], \tag{5·4}$$

where the first expression on the right denotes the conditional probability of the relation indicated for fixed values of the a_{irs}, and the second denotes the product of the probability elements of the n Wishart distributions of the a_{irs}. Now

$$\begin{aligned}\Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(a) \leqslant 2h(a) \mid a] &= \exp\left[\sum_i \sum_{r \leqslant s} (a_{irs} - \alpha_{irs}) \frac{\partial}{\partial \alpha_{irs}}\right] \Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha) \leqslant 2h(\alpha)] \\ &= \exp[\Sigma \operatorname{tr}(\mathbf{A}_i - \boldsymbol{\alpha}_i)\,\boldsymbol{\partial}_i] \Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha) \leqslant 2h(\alpha)],\end{aligned} \tag{5·5}$$

where $\boldsymbol{\partial}_i$ denotes the matrix of derivative operators:

$$\boldsymbol{\partial}_i = \begin{bmatrix} \partial/\partial\alpha_{i11} & \ldots & \frac{1}{2}\partial/\partial\alpha_{i1p} \\ \ldots & \ldots & \ldots \\ \frac{1}{2}\partial/\partial\alpha_{ip1} & \ldots & \partial/\partial\alpha_{ipp} \end{bmatrix}, \tag{5·6}$$

its typical element being
$$\partial_{irs} = \tfrac{1}{2}(1 + \delta_{rs})\,\partial/\partial\alpha_{irs}. \tag{5·7}$$

Thus (5·3) becomes
$$\Theta \Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha) \leqslant 2h(\alpha)] = G_\rho(\xi), \tag{5·8}$$

where
$$\Theta = \int \exp[\Sigma \operatorname{tr}(\mathbf{A}_i - \boldsymbol{\alpha}_i)\,\boldsymbol{\partial}_i] \Pr[da]. \tag{5·9}$$

We now see that Θ is $\exp[-\Sigma \operatorname{tr} \boldsymbol{\alpha}_i \boldsymbol{\partial}_i]$ times the product of the moment-generating functions of the n Wishart distributions (the 'variables' being ∂_{irs}) or

$$\Theta = \exp[-\Sigma \operatorname{tr} \boldsymbol{\alpha}_i \boldsymbol{\partial}_i]\, \Pi \left| \mathbf{I} - \frac{2}{\nu_i} \boldsymbol{\alpha}_i \boldsymbol{\partial}_i \right|^{-\frac{1}{2}\nu_i}. \tag{5·10}$$

To proceed, we require to prove the theorem enunciated in § 2, namely,

$$-\log|\mathbf{I} - \mathbf{Y}| = \operatorname{tr}\mathbf{Y} + \tfrac{1}{2}\operatorname{tr}\mathbf{Y}^2 + \tfrac{1}{3}\operatorname{tr}\mathbf{Y}^3 + \ldots \tag{5·11}$$

(the series being convergent, in the case when $\mathbf{Y}$ is a numerical matrix, when all the eigenvalues of $\mathbf{Y}$ are numerically less than unity). Suppose that the eigenvalues of $\mathbf{Y}$ are $\lambda_1, \ldots, \lambda_p$. It can be shown that the eigenvalues of $\mathbf{Y}^r$ are $\lambda_1^r, \ldots, \lambda_p^r$, for $r = 1, 2, \ldots$. Thus

$$|\mu\mathbf{I} - \mathbf{Y}^r| = (\mu - \lambda_1^r) \ldots (\mu - \lambda_p^r) \quad (r = 1, 2, \ldots). \tag{5·12}$$

Therefore
$$\operatorname{tr}\mathbf{Y}^r = \Sigma\lambda_i^r. \tag{5·13}$$

Also, using (5·12) with $r = 1, \mu = 1$,

$$\begin{aligned} -\log|\mathbf{I}-\mathbf{Y}| &= -\Sigma\log(1-\lambda_i) \\ &= \Sigma(\lambda_i+\tfrac{1}{2}\lambda_i^2+\tfrac{1}{3}\lambda_i^3+\ldots) \\ &= \operatorname{tr}\mathbf{Y}+\tfrac{1}{2}\operatorname{tr}\mathbf{Y}^2+\tfrac{1}{3}\operatorname{tr}\mathbf{Y}^3+\ldots, \end{aligned} \tag{5·14}$$

by (5·13), which proves the theorem.

Applying this result to (5·10) we obtain

$$\begin{aligned} \Theta &= \exp\Sigma\left[-\operatorname{tr}\boldsymbol{\alpha}_i\boldsymbol{\partial}_i-\tfrac{1}{2}\nu_i\log\left|\mathbf{I}-\frac{2}{\nu_i}\boldsymbol{\alpha}_i\boldsymbol{\partial}_i\right|\right] \\ &= \exp\Sigma\left[\frac{\operatorname{tr}(\boldsymbol{\alpha}_i\boldsymbol{\partial}_i)^2}{\nu_i}+\frac{4}{3}\frac{\operatorname{tr}(\boldsymbol{\alpha}_i\boldsymbol{\partial}_i)^3}{\nu_i^2}+\ldots\right] \\ &= 1+\Sigma\frac{\operatorname{tr}(\boldsymbol{\alpha}_i\boldsymbol{\partial}_i)^2}{\nu_i}+\left\{\tfrac{4}{3}\Sigma\frac{\operatorname{tr}(\boldsymbol{\alpha}_i\boldsymbol{\partial}_i)^3}{\nu_i^2}+\frac{1}{2}\left(\Sigma\frac{\operatorname{tr}(\boldsymbol{\alpha}_i\boldsymbol{\partial}_i)^2}{\nu_i}\right)^2\right\}+O(\nu^{-3}). \end{aligned} \tag{5·15}$$

The analogy with Welch's Θ operator is obvious. It should be noted that the operators $\boldsymbol{\partial}_i$ do not act on the $\boldsymbol{\alpha}_i$ present in Θ itself, and indeed it is more useful to write (5·15) in suffix form:

$$\begin{aligned} \Theta = 1&+\sum_i\sum_{rstu}\frac{\alpha_{iur}\alpha_{ist}\partial_{irs}\partial_{itu}}{\nu_i} \\ &+\left\{\tfrac{4}{3}\Sigma\Sigma\frac{\alpha_{iur}\alpha_{ist}\alpha_{ivu}\partial_{irs}\partial_{itu}\partial_{ivw}}{\nu_i^2}+\tfrac{1}{2}\Sigma\Sigma\frac{\alpha_{iur}\alpha_{ist}\alpha_{jyv}\alpha_{jwx}\partial_{irs}\partial_{itu}\partial_{jvw}\partial_{jxy}}{\nu_i\nu_j}\right\}+O(\nu^{-3}). \end{aligned} \tag{5·16}$$

Proceeding from (5·8) and (5·16) as in the univariate case we obtain, for the equations giving $h_1(\alpha)$ and $h_2(\alpha)$,

$$\left[h_1(\alpha)\,D+\Sigma\Sigma\frac{\alpha_{iur}\alpha_{ist}\partial_{irs}\partial_{itu}}{\nu_i}\right]\Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha)\leqslant 2\xi]=0, \tag{5·17}$$

$$\begin{aligned} \Big[h_2(\alpha)\,D+\tfrac{1}{2}h_1^2(\alpha)\,D^2+\Sigma\Sigma\frac{\alpha_{iur}\alpha_{ist}}{\nu_i}&(h_1^{(irs,itu)}(\alpha)\,D+2h_1^{(irs)}(\alpha)\,\partial_{itu}D \\ &+h_1(\alpha)\,\partial_{irs}\partial_{itu}D)+\tfrac{4}{3}\Sigma\Sigma\frac{\alpha_{iur}\alpha_{ist}\alpha_{ivu}\partial_{irs}\partial_{itu}\partial_{ivw}}{\nu_i^2} \\ &+\tfrac{1}{2}\Sigma\Sigma\frac{\alpha_{iur}\alpha_{ist}\alpha_{jyv}\alpha_{jwx}\partial_{irs}\partial_{itu}\partial_{jvw}\partial_{jxy}}{\nu_i\nu_j}\Big]\Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha)\leqslant 2\xi]=0, \end{aligned} \tag{5·18}$$

where

$$h_1^{(irs)}(\alpha)=\partial_{irs}h_1(\alpha)=\tfrac{1}{2}(1+\delta_{rs})\,\partial h_1(\alpha)/\partial\alpha_{irs}, \quad \text{etc.} \tag{5·19}$$

To evaluate the derivatives we consider

$$J=\Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha+\epsilon)\leqslant 2\xi], \tag{5·20}$$

where the increments $\boldsymbol{\epsilon}_1,\ldots,\boldsymbol{\epsilon}_n$ to $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n$ are symmetric matrices. As in equation (2·17) this is equivalent to

$$J=[1+\Sigma\Sigma\epsilon_{irs}\partial_{irs}+\tfrac{1}{2}\Sigma\Sigma\epsilon_{irs}\epsilon_{jtu}\partial_{irs}\partial_{jtu}+\ldots]\Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha)\leqslant 2\xi], \tag{5·21}$$

the sums being over $i,j,\ldots=1,\ldots,n;\ r,s,\ldots=1,\ldots,p$. But we can also express J in the form (2·18), and carrying out similar transformations and expansions on this to the ones used in § 2 we find, in the *simple case* when $\mathbf{t}(\alpha)$ is not actually a function of the α_{irs},

$$\begin{aligned} J=[1+\Sigma\Sigma\epsilon_{irs}&\{-\tfrac{1}{2}(irs)\,\Delta\} \\ &+\Sigma\Sigma\epsilon_{irs}\epsilon_{jtu}\{(irs\,|\,jtu)\,(\tfrac{1}{2}\Delta+\tfrac{1}{4}\Delta^2)+\tfrac{1}{8}(irs)\,(jtu)\,\Delta^2\}+\ldots]\,G_\rho(\xi), \end{aligned} \tag{5·22}$$

where $(irs) = \operatorname{tr} \mathbf{V}^{-1}\mathbf{V}_{irs}$, $(irs, jtu) = \operatorname{tr} \mathbf{V}^{-1}\mathbf{V}_{irs,jtu}$, $(irs \mid jtu) = \operatorname{tr} \mathbf{V}^{-1}\mathbf{V}_{irs}\mathbf{V}^{-1}\mathbf{V}_{jtu}, \ldots$ (5·23)

and $$\mathbf{V}_{irs} = \partial_{irs}\mathbf{V} = \tfrac{1}{2}(1+\delta_{rs})\,\partial\mathbf{V}/\partial\alpha_{irs}, \quad \mathbf{V}_{irs,jtu} = \partial_{irs}\partial_{jtu}\mathbf{V}, \ldots. \tag{5·24}$$

Thus we find $$\partial_{irs} \Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha) \leqslant 2\xi] \doteq \tfrac{1}{2}(irs)\,Eg_p(\xi), \tag{5·25}$$

$$\partial_{irs}\partial_{jtu} \Pr[\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}(\alpha) \leqslant 2\xi] = [-\tfrac{1}{2}(irs \mid jtu)\,(E^2+E) - \tfrac{1}{4}(irs)\,(jtu)\,(E^2-E)]\,g_p(\xi), \tag{5·26}$$

and similar modifications of (2·41) and (2·42). The rest of the work is similar to that in § 2, and the final result is

$$\begin{aligned}
2h(a) = \chi^2 &+ \tfrac{1}{2}\sum_i \sum_{rstu} \frac{a_{iur}a_{ist}}{\nu_i}[2(\chi_4+\chi_2)\,(irs \mid itu) + (\chi_4-\chi_2)\,(irs)\,(itu)] \\
&+ \left\{\frac{1}{16}\left(1-\frac{r-2}{\chi^2}\right)\left[\sum_i \sum_{rstu} \frac{a_{iur}a_{ist}}{\nu_i}[2(\chi_4+\chi_2)\,(irs \mid itu) + (\chi_4-\chi_2)\,(irs)\,(itu)]\right]^2\right. \\
&- \tfrac{1}{2}\sum_i \sum_{rstu} \frac{a_{iur}a_{ist}}{\nu_i^2}[2(\chi_4+\chi_2)\,((iur \mid ist) + (irs \mid itu)) + (\chi_4-\chi_2)\,((iur)\,(ist) + (irs)\,(itu))] \\
&+ 4\sum_i \sum_{\substack{rst\\uvw}} \frac{a_{iur}a_{ist}a_{iuv}}{\nu_i^2}[2(\chi_4+\chi_2)\,(irs \mid itu \mid ivw) + (\chi_4-\chi_2)\,(irs \mid itu)\,(ivw)] \\
&- \sum_{ij} \sum_{\substack{rstu\\vwxy}} \frac{a_{iur}a_{ist}a_{jyv}a_{jwx}}{\nu_i\nu_j}[2(\chi_4+\chi_2)\,(2(irs \mid itu \mid jvw \mid jxy) + (irs \mid jvw \mid itu \mid jxy)) \\
&\qquad + (\chi_4-\chi_2)\,(2(irs \mid itu \mid jvw)\,(jxy) + (irs \mid jvw)\,(itu \mid jxy))] \\
&+ (\chi_2-1)\sum_i \sum_{\substack{rst\\uvw}} \frac{a_{iur}a_{ist}a_{iuv}}{\nu_i^2}[2(\chi_4+\chi_2)\,(irs \mid itu)\,(ivw) + (\chi_4-\chi_2)\,(irs)\,(itu)\,(ivw)] \\
&- (\chi_2-1)\sum_{ij} \sum_{\substack{rstu\\vwxy}} \frac{a_{iur}a_{ist}a_{jyv}a_{jwx}}{\nu_i\nu_j}[2(\chi_4+\chi_2)\,(irs \mid itu \mid jvw)\,(jxy) \\
&\qquad + (\chi_4-\chi_2)\,(irs \mid jvw)\,(itu)\,(jxy)] \\
&- \tfrac{1}{8}\sum_{ij} \sum_{\substack{rstu\\vwxy}} \frac{a_{iur}a_{ist}a_{jyv}a_{jwx}}{\nu_i\nu_j}[2(\chi_4-1)\,(irs \mid itu) + (\chi_4-2\chi_2+1)\,(irs)\,(itu)] \\
&\qquad \times [2(\chi_4+\chi_2)\,(jvw \mid jxy) + (\chi_4-\chi_2)\,(jvw)\,(jxy)] \\
&- \tfrac{1}{3}\sum_i \sum_{\substack{rst\\uvw}} \frac{a_{iur}a_{ist}a_{iuv}}{\nu_i^2}[8(\chi_6+\chi_4+\chi_2)\,(irs \mid itu \mid ivw) + 6(\chi_6-\chi_2)\,(irs \mid itu)\,(ivw) \\
&\qquad + (\chi_6-2\chi_4+\chi_2)\,(irs)\,(itu)\,(ivw)] \\
&+ \tfrac{1}{16}\sum_{ij} \sum_{\substack{rstu\\vwxy}} \frac{a_{iur}a_{ist}a_{jyv}a_{jwx}}{\nu_i\nu_j} \\
&\times [16(\chi_8+\chi_6+\chi_4+\chi_2)\,(2(irs \mid itu \mid jvw \mid jxy) + (irs \mid jvw \mid itu \mid jxy)) \\
&\quad + 32(\chi_8-\chi_2)\,(irs \mid itu \mid jvw)\,(jxy) \\
&\quad + 4(\chi_8+\chi_6-\chi_4-\chi_2)\,((irs \mid itu)\,(jvw \mid jxy) + 2(irs \mid jvw)\,(itu \mid jxy)) \\
&\quad + 4(\chi_8-\chi_6-\chi_4+\chi_2)\,((irs \mid itu)\,(jvw)\,(jxy) + 2(irs \mid jvw)\,(itu)\,(jxy)) \\
&\quad \left. + (\chi_8-3\chi_6+3\chi_4-\chi_2)\,(irs)\,(itu)\,(jvw)\,(jxy)]\right\} + O(\nu^{-3}).
\end{aligned} \tag{5·27}$$

Here χ_{2s} denotes $\chi^{2s}/r(r+2)\ldots(r+2s-2)$ while (irs) now denotes $\operatorname{tr} \mathbf{V}^{-1}(a)\,\mathbf{V}_{irs}(a)$, where $\mathbf{V}_{irs}(a) = \tfrac{1}{2}(1+\delta_{rs})\,\partial\mathbf{V}(a)/\partial a_{irs}$, and so on.

In the *general case*, where $\mathbf{t}(\alpha)$ is actually a function of the α_{irs}, a very similar analysis to that given in § 2 finally yields

$$2h(a) = \chi^2 + (p+1)\chi_2 \sum_i \sum_{rs} \frac{a_{irs}}{\nu_i} \operatorname{tr} \mathbf{\Lambda}^{-1}\mathbf{\Lambda}_{irs}$$
$$+ \sum_i \sum_{rstu} \frac{a_{iur}a_{ist}}{\nu_i} [3\chi_2 \operatorname{tr} \mathbf{\Lambda}^{-1}\mathbf{\Lambda}_{irs,itu} + (\chi_4 - 5\chi_2) \operatorname{tr} \mathbf{\Lambda}^{-1}\mathbf{\Lambda}_{irs}\mathbf{\Lambda}^{-1}\mathbf{\Lambda}_{itu}$$
$$+ \tfrac{1}{2}(\chi_4 - \chi_2)(\operatorname{tr} \mathbf{\Lambda}^{-1}\mathbf{\Lambda}_{irs})(\operatorname{tr} \mathbf{\Lambda}^{-1}\mathbf{\Lambda}_{itu}) + 2\chi_2 \operatorname{tr} \mathbf{\Lambda}(\mathbf{C}^{-1}\mathbf{C}_{irs}\mathbf{C}^{-1}\mathbf{C}_{itu}\mathbf{C}^{-1})_{11}] + O(\nu^{-2}), \quad (5{\cdot}28)$$

the analogue of (2·59). ($\mathbf{\Lambda}$ denotes $\mathbf{V}^{-1}$.)

An immediate generalization of the above results is to the case when the vector $\mathbf{x}_i$ contains p_i variates, the p_i not being necessarily equal; apart from altering the (implicit) limits on summation signs, the work remains virtually unchanged.

Another generalization is analogous to that given at the end of § 2; in place of $\mathbf{var}\, \mathbf{x}_i = \boldsymbol{\alpha}_i$ we may write

$$\mathbf{var}\, \mathbf{x}_i = \sum_{j=1}^{m} \beta_{ij}\boldsymbol{\alpha}_j, \quad \mathbf{cov}\,(\mathbf{x}_i, \mathbf{x}_j) = 0 \quad (i \neq j). \quad (5{\cdot}29)$$

Both generalizations may be covered by writing

$$\operatorname{cov}(x_{ir}, x_{is}) = \sum_{j=1}^{m} \sum_{t,u=1}^{p_j} \beta_{irs}^{jtu}\alpha_{jtu}, \quad \operatorname{cov}(x_{ir}, x_{js}) = 0 \quad (i \neq j). \quad (5{\cdot}30)$$

6. The multivariate analogue of the Behrens-Fisher problem

Suppose that we have samples from two p-variate normal populations, whose true centroids are $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and whose variance matrices are unknown and not assumed to be equal. Let $\mathbf{x}_1$, $\mathbf{x}_2$ denote the centroids of the samples, and let the sampling variance matrices of these centroids be $\boldsymbol{\alpha}_1 = (\alpha_{1rs})$, $\boldsymbol{\alpha}_2 = (\alpha_{2rs})$, while $\mathbf{A}_1 = (a_{1rs})$, $\mathbf{A}_2 = (a_{2rs})$ are the usual unbiased estimates of $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$, distributed in Wishart's forms with ν_1, ν_2 degrees of freedom respectively. We wish to test the hypothesis $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$; this becomes $\boldsymbol{\theta} = 0$ if we write $\boldsymbol{\theta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. The maximum-likelihood estimate of $\boldsymbol{\theta}$ is $\mathbf{t} = \mathbf{x}_1 - \mathbf{x}_2$, which is independent of the unknown variance matrices, and its sampling variance matrix is $\mathbf{V}(\alpha) = \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$. Thus if $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ were known we could test our hypothesis by comparing $\mathbf{t}'\mathbf{V}^{-1}(\alpha)\,\mathbf{t}$ with a significance point of χ^2 on p degrees of freedom, but as they are not we compare $\mathbf{t}'\mathbf{V}^{-1}(a)\,\mathbf{t} \equiv \mathbf{t}'\mathbf{V}^{-1}\mathbf{t}$ with $2h(a)$ given by (5·27).

Now $\mathbf{V}_{irs}$ is a matrix with zeros everywhere except for $\frac{1}{2}$'s in the places (r,s) and (s,r) (or 1 in the place (r,r) if $r = s$). Thus

$$(irs) = \operatorname{tr} \mathbf{V}^{-1}\mathbf{V}_{irs} = \sum_{ab} v^{ba}\, \tfrac{1}{2}(\delta_{ra}\delta_{sb} + \delta_{sa}\delta_{rb})$$
$$= \tfrac{1}{2}(v^{sr} + v^{rs}) = v^{sr}, \quad (6{\cdot}1)$$

$$(irs \mid jtu) = \sum_{abcd} v^{da}\, \tfrac{1}{2}(\delta_{ra}\delta_{sb} + \delta_{sa}\delta_{rb})\, v^{bc}\, \tfrac{1}{2}(\delta_{tc}\delta_{ud} + \delta_{uc}\delta_{td})$$
$$= \tfrac{1}{4}(v^{ur}v^{st} + v^{us}v^{rt} + v^{tr}v^{su} + v^{ts}v^{ru})$$
$$= \tfrac{1}{2}(v^{ur}v^{st} + v^{us}v^{rt}). \quad (6{\cdot}2)$$

(The final reduction to half the number of terms is a peculiarity not shared by traces of higher order; apart from this it will be seen that the totality of the terms are obtained

from the first by interchanging the letter pairs r, s; t, u; ... occurring together on the left-hand sides.) In this fashion we find

$$\sum_{rstu} a_{iur}a_{ist}(irs \mid itu) = \tfrac{1}{2}([i \mid i] + [i]^2), \tag{6·3}$$

$$\sum_{rstu} a_{iur}a_{ist}(irs)(itu) = [i \mid i], \tag{6·4}$$

where $$[i] = \operatorname{tr} \mathbf{V}^{-1}\mathbf{A}_i, \quad [i \mid j] = \operatorname{tr} \mathbf{V}^{-1}\mathbf{A}_i\mathbf{V}^{-1}\mathbf{A}_j, \ldots, \quad \text{and} \quad \mathbf{V} = \Sigma \mathbf{A}_i. \tag{6·5}$$

Hence $$2h(a) = \chi^2 + \tfrac{1}{2}\Sigma \frac{1}{\nu_i}(2\chi_4[i \mid i] + (\chi_4 + \chi_2)[i]^2) + O(\nu^{-2}). \tag{6·6}$$

Here χ^2 is based on p degrees of freedom, and χ_{2s} denotes $\chi^{2s}/p(p+2)\ldots(p+2s-2)$. The term of order ν^{-2}, obtained from further formulae like (6·3) and (6·4), over which a good deal of care is necessary, is

$$\begin{aligned}
2h_2(a) = {} & \frac{1}{16}\left(1 - \frac{p-2}{\chi^2}\right)\left[\Sigma \frac{1}{\nu_i}(2\chi_4[i \mid i] + (\chi_4 + \chi_2)[i]^2)\right]^2 \\
& - \Sigma \nu_i^{-2}((2\chi_4 + \chi_2)[i \mid i] + \chi_4[i]^2) \\
& + \Sigma \nu_i^{-2}(2(3\chi_4 + \chi_2)[i \mid i \mid i] + (5\chi_4 + \chi_2)[i \mid i][i] + (\chi_4 + \chi_2)[i]^3) \\
& - \Sigma (\nu_i\nu_j)^{-1}(2\chi_4[i \mid i \mid j \mid j] + (2\chi_4 + \chi_2)[i \mid j \mid i \mid j] + (3\chi_4 + \chi_2)[i \mid i \mid j][j] \\
& \quad + \chi_4[i \mid j]^2 + (\chi_4 + \chi_2)[i \mid j][i][j]) \\
& + (\chi_2 - 1)\Sigma \nu_i^{-2}(2\chi_4[i \mid i \mid i] + (\chi_4 + \chi_2)[i \mid i][i]) \\
& - (\chi_2 - 1)\Sigma (\nu_i\nu_j)^{-1}(2\chi_4[i \mid i \mid j \mid j] + (\chi_4 + \chi_2)[i \mid i \mid j][j]) \\
& - \tfrac{1}{8}\Sigma (\nu_i\nu_j)^{-1}(2(\chi_4 - \chi_2)[i \mid i] + (\chi_4 - 1)[i]^2)(2\chi_4[j \mid j] + (\chi_4 + \chi_2)[j]^2) \\
& - \tfrac{1}{3}\Sigma \nu_i^{-2}(2(4\chi_6 + \chi_4 + \chi_2)[i \mid i \mid i] + 3(2\chi_6 + \chi_4)[i \mid i][i] + (\chi_6 + \chi_4 + \chi_2)[i]^3) \\
& + \tfrac{1}{16}\Sigma (\nu_i\nu_j)^{-1}(32\chi_8[i \mid i \mid j \mid j] + 8(2\chi_8 + 2\chi_6 + \chi_4 + \chi_2)[i \mid j \mid i \mid j] \\
& \quad + 16(2\chi_8 + \chi_6 + \chi_4)[i \mid i \mid j][j] + 4(\chi_8 - \chi_6)[i \mid i][j \mid j] \\
& \quad + 8(\chi_8 + \chi_6)[i \mid j]^2 + 4(\chi_8 - \chi_4)[i \mid i][j]^2 \\
& \quad + 8(\chi_8 + \chi_6 + \chi_4 + \chi_2)[i \mid j][i][j] + (\chi_8 + \chi_6 - \chi_4 - \chi_2)[i]^2[j]^2).
\end{aligned} \tag{6·7}$$

This result has been checked by treating the problem on its own merits, and not as an example of the general linear hypothesis; however this method seems to the author to be at least as confusing as the one adopted here.

In the univariate case the result reduces to

$$\begin{aligned}
2h(a) = \chi^2\Bigg[1 + \tfrac{1}{2}(\chi^2 + 1)\frac{\Sigma a_i^2/\nu_i}{(\Sigma a_i)^2} + \Bigg\{ & -(\chi^2 + 1)\frac{\Sigma a_i^2/\nu_i^2}{(\Sigma a_i)^2} + \tfrac{2}{3}(\chi^4 + 5\chi^2 + 3)\frac{\Sigma a_i^3/\nu_i^2}{(\Sigma a_i)^3} \\
& - \tfrac{1}{8}(4\chi^4 + 15\chi^2 + 7)\frac{(\Sigma a_i^2/\nu_i)^2}{(\Sigma a_i)^4}\Bigg\}\Bigg] + O(\nu^{-3}),
\end{aligned} \tag{6·8}$$

which is equivalent to Welch's (1947*a*) result.

Of course there is nothing to stop us considering, with Welch, the slightly more general problem where $\mathbf{y}$ is a vector having a p-variate normal distribution with centroid $\boldsymbol{\eta}$ and variance matrix† $\mathbf{V}(\alpha) = \boldsymbol{\alpha}_1 + \ldots + \boldsymbol{\alpha}_k$, while $\mathbf{A}_1, \ldots, \mathbf{A}_k$ are distributed independently in Wishart distributions with expectations $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k$ and degrees of freedom $\nu_1, \ldots, \nu_k$. The test for the hypothesis $\boldsymbol{\eta} = 0$ (say) is then to compare $\mathbf{y}'\mathbf{V}^{-1}(a)\,\mathbf{y}$ with $2h(a)$ given by (6·6) and (6·7), the sums therein now running from 1 to k instead of from 1 to 2.

† Or $\mathbf{V}(\alpha) = \Sigma \lambda_i \boldsymbol{\alpha}_i$; the inclusion or omission of constant multipliers is a trivial matter since they can always be absorbed into the $\boldsymbol{\alpha}_i$ and $\mathbf{A}_i$.

7. A TEST FOR THE EQUALITY OF THE CENTROIDS OF SEVERAL MULTIVARIATE POPULATIONS

We now extend the work of the preceding section to the case of samples from k populations. Using the same notation as before, we write

$$\mathscr{E}(\mathbf{x}_i) = \boldsymbol{\theta}_i + \boldsymbol{\theta}_k \quad (i = 1, \ldots, k-1), \tag{7·1}$$

$$\mathscr{E}(\mathbf{x}_k) = \boldsymbol{\theta}_k, \tag{7·2}$$

the $\mathbf{x}_i$ being the sample centroids, and the hypothesis to be tested being that

$$\boldsymbol{\theta}_1 = \ldots = \boldsymbol{\theta}_{k-1} = 0.$$

The maximum-likelihood solution gives us

$$\mathbf{t}_i = \mathbf{x}_i - \mathbf{x}_k \quad (i = 1, \ldots, k-1). \tag{7·3}$$

Denoting $\{\mathbf{t}_1, \ldots, \mathbf{t}_{k-1}\}$ by $\mathbf{t}$, its estimated variance matrix is

$$\mathbf{V} = \begin{bmatrix} \mathbf{A}_1 + \mathbf{A}_k & \mathbf{A}_k & \ldots & \mathbf{A}_k \\ \mathbf{A}_k & \mathbf{A}_2 + \mathbf{A}_k & \ldots & \mathbf{A}_k \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{A}_k & \mathbf{A}_k & \ldots & \mathbf{A}_{k-1} + \mathbf{A}_k \end{bmatrix}, \tag{7·4}$$

and

$$\mathbf{V}^{-1} = \begin{bmatrix} \mathbf{W}_1 - \mathbf{W}_1\mathbf{W}^{-1}\mathbf{W}_1 & -\mathbf{W}_1\mathbf{W}^{-1}\mathbf{W}_2 & \ldots & -\mathbf{W}_1\mathbf{W}^{-1}\mathbf{W}_{k-1} \\ -\mathbf{W}_2\mathbf{W}^{-1}\mathbf{W}_1 & \mathbf{W}_2 - \mathbf{W}_2\mathbf{W}^{-1}\mathbf{W}_2 & \ldots & -\mathbf{W}_2\mathbf{W}^{-1}\mathbf{W}_{k-1} \\ \ldots & \ldots & \ldots & \ldots \\ -\mathbf{W}_{k-1}\mathbf{W}^{-1}\mathbf{W}_1 & -\mathbf{W}_{k-1}\mathbf{W}^{-1}\mathbf{W}_2 & \ldots & \mathbf{W}_{k-1} - \mathbf{W}_{k-1}\mathbf{W}^{-1}\mathbf{W}_{k-1} \end{bmatrix}, \tag{7·5}$$

where the $\mathbf{A}_i$ are the estimates of $\alpha_i = \mathbf{var}\,\mathbf{x}_i$ based on ν_i degrees of freedom and where the 'weight matrices' $\mathbf{W}_i$ and $\mathbf{W}$ are defined by

$$\mathbf{W}_i = \mathbf{A}_i^{-1}, \quad \mathbf{W} = \sum_1^k \mathbf{W}_i. \tag{7·6}$$

As for the derivatives, if $i \neq k$ we see that $\mathbf{V}_{irs}$ is a matrix with zeros everywhere except for $\frac{1}{2}$'s in the places (ir, is) and (is, ir), or, if $r = s$, except for unity in the place (ir, ir); on the other hand, $\mathbf{V}_{krs}$ has all its elements zero except for $\frac{1}{2}$'s in the places (ir, js) and (is, jr), for all $i, j = 1, \ldots, k-1$, or, if $r = s$, except for units in all the places (ir, jr). Thus (compare equations (6·1) and (6·2))

$$(irs) = v^{is, ir}, \tag{7·7}$$

$$(irs \mid jtu) = \tfrac{1}{2}(v^{ju, ir}v^{is, jt} + v^{ju, is}v^{ir, jt}), \tag{7·8}$$

where $v^{ir, js}$ denotes the (ir, js)th element of $\mathbf{V}^{-1}$. Strictly speaking these formulae only apply when $i, j \neq k$, but if (see (7·5)) we interpret $v^{ir, js}$ as the (r, s)th element of

$$(\delta_{ij}\mathbf{W}_i - \mathbf{W}_i\mathbf{W}^{-1}\mathbf{W}_j),$$

even when i or j is equal to k, they can be shown to hold universally (by the same kind of appeal to the invariance of traces under linear transformations of the t's as was made in the univariate case). Hence, substituting these results into the *simple case* formula (5·27) we obtain

$$2h(a) = \chi^2 + \tfrac{1}{2}\sum_i \sum_{rstu} \frac{a_{iur}a_{ist}}{\nu_i}((\chi_4 + \chi_2)(v^{iu, ir}v^{is, it} + v^{iu, is}v^{ir, it}) + (\chi_4 - \chi_2)v^{is, ir}v^{iu, it}) + O(\nu^{-2})$$

$$= \chi^2 + \tfrac{1}{2}\Sigma\frac{1}{\nu_i}(2\chi_4 \operatorname{tr}(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i)^2 + (\chi_4 + \chi_2)[\operatorname{tr}(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i)]^2) + O(\nu^{-2}), \tag{7·9}$$

where χ^2 is based on $r = p(k-1)$ degrees of freedom and χ_{2s} denotes $\chi^{2s}/r(r+2)\dots(r+2s-2)$. The term of order ν^{-2} is given by the same formula (6·7) that applies in the problem of two centroids, provided that we reinterpret the trace symbols as follows:

$$[i \mid j] = \operatorname{tr}(\delta_{ji}\mathbf{W}_j - \mathbf{W}_j\mathbf{W}^{-1}\mathbf{W}_i)\,\mathbf{A}_i(\delta_{ij}\mathbf{W}_i - \mathbf{W}_i\mathbf{W}^{-1}\mathbf{W}_j)\,\mathbf{A}_j$$

$$= \operatorname{tr}(\delta_{ji}\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i)(\delta_{ij}\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_j) \tag{7·10}$$

and so on. These reduce to the former expressions $\operatorname{tr}\mathbf{V}^{-1}\mathbf{A}_i\mathbf{V}^{-1}\mathbf{A}_j$ etc., in the case $k = 2$. The full set of formulae necessary to interpret (6·7) is

$$[i] = \operatorname{tr}(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i), \tag{7·11}$$

$$[i \mid i] = \operatorname{tr}(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i)^2, \tag{7·12}$$

$$[i \mid j] = \operatorname{tr}(\delta_{ij}\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i)(\delta_{ij}\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_j), \tag{7·13}$$

$$[i \mid i \mid i] = \operatorname{tr}(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i)^3, \tag{7·14}$$

$$[i \mid i \mid j] = \operatorname{tr}(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i)(\delta_{ij}\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_j)(\delta_{ij}\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i), \tag{7·15}$$

$$[i \mid i \mid j \mid j] = \operatorname{tr}(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i)(\delta_{ij}\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_j)(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_j)(\delta_{ij}\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i), \tag{7·16}$$

$$[i \mid j \mid i \mid j] = \operatorname{tr}[(\delta_{ij}\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i)(\delta_{ij}\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_j)]^2. \tag{7·17}$$

The result could obviously be written out explicitly in the form of the equation (3·30), but as the expression would be lengthier, and involve a more extensive set of R symbols, it is not proposed to do so here.

The quantity which has to be compared with $2h(a)$ is $\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}$, which, from equations (7·3) and (7·5), is given by

$$\mathbf{t}'\mathbf{V}^{-1}\mathbf{t} = \Sigma\mathbf{x}_i'\mathbf{W}_i\mathbf{x}_i - \Sigma\mathbf{x}_i'\mathbf{W}_i\mathbf{W}^{-1}\mathbf{W}_j\mathbf{x}_j$$

$$= \Sigma(\mathbf{x}_i - \bar{\mathbf{x}})'\,\mathbf{W}_i(\mathbf{x}_i - \bar{\mathbf{x}}), \tag{7·18}$$

provided that we define

$$\bar{\mathbf{x}} = \mathbf{W}^{-1}\Sigma\mathbf{W}_i\mathbf{x}_i. \tag{7·19}$$

This is the analogue of the weighted sum of squares $\Sigma w_i x_i^2 - (\Sigma w_i x_i)^2/w = \Sigma w_i(x_i - \bar{x})^2$ occurring in the corresponding univariate problem. On large-sample theory it would be compared with χ^2 based on $r = p(k-1)$ degrees of freedom, and equations (7·9), and (6·7) interpreted with the help of (7·11)–(7·17) give further approximations.

8. Numerical illustrations of multivariate tests

Suppose that we have three samples, from bivariate normal populations, of sizes 16, 11, 11, and that their means or centroids are

$$\mathbf{x}_1 = \begin{bmatrix} 9{\cdot}82 \\ 15{\cdot}06 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 13{\cdot}05 \\ 22{\cdot}57 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 14{\cdot}67 \\ 25{\cdot}17 \end{bmatrix}, \tag{8·1}$$

and their variance matrices (using divisors $\nu_1 = 15$, $\nu_2 = 10$, $\nu_3 = 10$) are

$$\begin{bmatrix} 120{\cdot}0 & -16{\cdot}3 \\ -16{\cdot}3 & 17{\cdot}8 \end{bmatrix}, \quad \begin{bmatrix} 81{\cdot}8 & 32{\cdot}1 \\ 32{\cdot}1 & 53{\cdot}8 \end{bmatrix}, \quad \begin{bmatrix} 100{\cdot}3 & 23{\cdot}2 \\ 23{\cdot}2 & 97{\cdot}1 \end{bmatrix}. \tag{8·2}$$

The estimated variance matrices of the means $\mathbf{x}_i$, obtained by dividing (8·2) by the sample sizes, are

$$\mathbf{A}_1 = \begin{bmatrix} 7{\cdot}500 & -1{\cdot}019 \\ -1{\cdot}019 & 1{\cdot}112 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 7{\cdot}436 & 2{\cdot}918 \\ 2{\cdot}918 & 4{\cdot}891 \end{bmatrix}, \quad \mathbf{A}_3 = \begin{bmatrix} 9{\cdot}118 & 2{\cdot}109 \\ 2{\cdot}109 & 8{\cdot}827 \end{bmatrix}. \tag{8·3}$$

Purely as a convenience, we shall utilize all three samples to illustrate the test of § 7, and the first two samples to illustrate that of § 6, but we shall not be concerning ourselves with questions arising from the non-independence of the two tests.

(a) Test for two centroids

In the notation of § 6 the difference between the two sample means is

$$\mathbf{t} = \mathbf{x}_1 - \mathbf{x}_2 = \begin{bmatrix} -3{\cdot}23 \\ -7{\cdot}51 \end{bmatrix}, \tag{8·4}$$

and its estimated variance matrix is

$$\mathbf{V} = \mathbf{A}_1 + \mathbf{A}_2 = \begin{bmatrix} 14{\cdot}936 & 1{\cdot}899 \\ 1{\cdot}899 & 6{\cdot}003 \end{bmatrix}. \tag{8·5}$$

Thus
$$\mathbf{t}'\mathbf{V}^{-1}\mathbf{t} = [-3{\cdot}23 \quad -7{\cdot}51] \begin{bmatrix} 0{\cdot}06976 & -0{\cdot}02207 \\ -0{\cdot}02207 & 0{\cdot}17357 \end{bmatrix} \begin{bmatrix} -3{\cdot}23 \\ -7{\cdot}51 \end{bmatrix} = 9{\cdot}45. \tag{8·6}$$

On large-sample theory this would merely be compared with the appropriate significance point of χ^2 with two degrees of freedom (corresponding to the two hypothetical restrictions, $\mu_{11} = \mu_{12}$, $\mu_{21} = \mu_{22}$). A selection of these significance points is given in the second column of Table 1, from which it will be seen that on this test the centroids would be judged to be significantly different at the 1 % level. However, the large-sample test does not take into account the fact that $\mathbf{A}_1$ and $\mathbf{A}_2$ are only *estimates* of the true variance matrices $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$, and so it tends to exaggerate the significance of the difference. The use of the corrected form (6·6) will at least give some idea of the effect of this 'Studentization'; Welch's experience with the univariate problem suggests that the first corrective term may account for most of the 'Student' correction, but in any case its use will at least guard us against drawing over-enthusiastic conclusions from the large sample test.

Table 1

Significance level (%)	χ^2	$A + B\chi^2$	$2h(a)$
5	5·991	1·207	7·23
2½	7·378	1·244	9·18
1	9·210	1·292	11·90

Now (6·6) may be written in the form

$$2h(a) = \chi^2(A + B\chi^2), \tag{8·7}$$

where
$$A = 1 + \frac{1}{2p}\Sigma\frac{(\operatorname{tr}\mathbf{V}^{-1}\mathbf{A}_i)^2}{\nu_i}, \tag{8·8}$$

$$B = \frac{1}{p(p+2)}\left[\Sigma\frac{\operatorname{tr}(\mathbf{V}^{-1}\mathbf{A}_i)^2}{\nu_i} + \tfrac{1}{2}\Sigma\frac{(\operatorname{tr}\mathbf{V}^{-1}\mathbf{A}_i)^2}{\nu_i}\right]. \tag{8·9}$$

Taking $\mathbf{V}^{-1}$ from (8·6) we find

$$\mathbf{V}^{-1}\mathbf{A}_1 = \begin{bmatrix} 0{\cdot}5457 & -0{\cdot}0956 \\ -0{\cdot}3424 & 0{\cdot}2155 \end{bmatrix}, \quad \mathbf{V}^{-1}\mathbf{A}_2 = \begin{bmatrix} 0{\cdot}4543 & 0{\cdot}0956 \\ 0{\cdot}3424 & 0{\cdot}7845 \end{bmatrix}; \tag{8·10}$$

$$\operatorname{tr}(\mathbf{V}^{-1}\mathbf{A}_1)^2 = 0{\cdot}4097, \qquad \operatorname{tr}(\mathbf{V}^{-1}\mathbf{A}_2)^2 = 0{\cdot}8873;\dagger \tag{8·11}$$

$$(\operatorname{tr}\mathbf{V}^{-1}\mathbf{A}_1)^2 = 0{\cdot}5794, \qquad (\operatorname{tr}\mathbf{V}^{-1}\mathbf{A}_2)^2 = 1{\cdot}5346; \tag{8·12}$$

$$\Sigma\frac{\operatorname{tr}(\mathbf{V}^{-1}\mathbf{A}_i)^2}{\nu_i} = 0{\cdot}11604, \quad \Sigma\frac{(\operatorname{tr}\mathbf{V}^{-1}\mathbf{A}_i)^2}{\nu_i} = 0{\cdot}19209; \tag{8·13}$$

whence

$$A = 1+\tfrac{1}{4}(0{\cdot}19209) = 1{\cdot}0480, \tag{8·14}$$

$$B = \tfrac{1}{8}[0{\cdot}11604+\tfrac{1}{2}(0{\cdot}19209)] = 0{\cdot}02651. \tag{8·15}$$

The values of the multiplier $A+B\chi^2$ and of $2h(a)$ (to the order ν^{-1}) are given in Table 1. On this basis our figure of 9·45 (equation (8·6)) can no longer be regarded as significant at the 1 % level, but it is so at the 2½ % level.

An approximate 95 % confidence region for the true centroid difference, $\boldsymbol{\delta} \equiv \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, is given by

$$(\mathbf{t}-\boldsymbol{\delta})'\,\mathbf{V}^{-1}(\mathbf{t}-\boldsymbol{\delta}) \leqslant 2h(a), \tag{8·16}$$

i.e.

$$0{\cdot}06976(\delta_1+3{\cdot}23)^2 - 0{\cdot}04414(\delta_1+3{\cdot}23)(\delta_2+7{\cdot}51) + 0{\cdot}17357(\delta_2+7{\cdot}51)^2 \leqslant 7{\cdot}23. \tag{8·17}$$

(b) Test for three centroids

In the notation of § 7 we have

$$\mathbf{W}_1 = \mathbf{A}_1^{-1} = \begin{bmatrix} 0{\cdot}1523 & 0{\cdot}1396 \\ 0{\cdot}1396 & 1{\cdot}0272 \end{bmatrix}, \quad \mathbf{W}_2 = \begin{bmatrix} 0{\cdot}1756 & -0{\cdot}1048 \\ -0{\cdot}1048 & 0{\cdot}2670 \end{bmatrix}, \quad \mathbf{W}_3 = \begin{bmatrix} 0{\cdot}1161 & -0{\cdot}0277 \\ -0{\cdot}0277 & 0{\cdot}1199 \end{bmatrix}; \tag{8·18}$$

$$\mathbf{W} = \Sigma\mathbf{W}_i = \begin{bmatrix} 0{\cdot}4440 & 0{\cdot}0071 \\ 0{\cdot}0071 & 1{\cdot}4141 \end{bmatrix}, \quad \mathbf{W}^{-1} = \begin{bmatrix} 2{\cdot}2524 & -0{\cdot}0113 \\ -0{\cdot}0113 & 0{\cdot}7072 \end{bmatrix}; \tag{8·19}$$

$$\mathbf{W}_1\mathbf{x}_1 = \begin{bmatrix} 3{\cdot}5980 \\ 16{\cdot}8405 \end{bmatrix}, \quad \mathbf{W}_2\mathbf{x}_2 = \begin{bmatrix} -0{\cdot}0738 \\ 4{\cdot}6586 \end{bmatrix}, \quad \mathbf{W}_3\mathbf{x}_3 = \begin{bmatrix} 1{\cdot}0060 \\ 2{\cdot}6115 \end{bmatrix}; \tag{8·20}$$

$$\Sigma\mathbf{W}_i\mathbf{x}_i = \mathbf{W}\bar{\mathbf{x}} = \begin{bmatrix} 4{\cdot}5302 \\ 24{\cdot}1106 \end{bmatrix}; \quad \bar{\mathbf{x}} = \mathbf{W}^{-1}\Sigma\mathbf{W}_i\mathbf{x}_i = \begin{bmatrix} 9{\cdot}9314 \\ 16{\cdot}9998 \end{bmatrix}. \tag{8·21}$$

Thus

$$\Sigma(\mathbf{x}_i-\bar{\mathbf{x}})'\,\mathbf{W}_i(\mathbf{x}_i-\bar{\mathbf{x}}) = \Sigma\mathbf{x}_i'(\mathbf{W}_i\mathbf{x}_i) - \bar{\mathbf{x}}'(\mathbf{W}\bar{\mathbf{x}}) = 18{\cdot}75. \tag{8·22}$$

If we were to use the large-sample test, this would be compared with significance points of χ^2 on four degrees of freedom (corresponding to the four hypothetical constraints $\mu_{11} = \mu_{12} = \mu_{13}, \mu_{21} = \mu_{22} = \mu_{23}$) as given in the second column of Table 2. On this (incorrect) basis the hypothesis would be regarded as disproved at the 0·1 % level.

Table 2

Significance level (%)	χ^2	$A+B\chi^2$	$2h(a)$
5	9·488	1·283	12·17
1	13·277	1·369	18·18
0·1	18·467	1·488	27·48

† Note that $\operatorname{tr}\mathbf{Y}^2 = \Sigma_{ij}y_{ij}y_{ji} = \Sigma_i y_{ii}^2 + 2\Sigma_{i<j}y_{ij}y_{ji}$.

The effect of 'Studentization' may be accounted for, to the order ν^{-1}, by the use of (7·9), which may be written

$$2h(a) = \chi^2(A + B\chi^2), \tag{8·23}$$

where

$$A = 1 + \frac{1}{2r}\Sigma\frac{[\mathrm{tr}\,(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i)]^2}{\nu_i}, \tag{8·24}$$

$$B = \frac{1}{r(r+2)}\left[\Sigma\frac{\mathrm{tr}\,(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i)^2}{\nu_i} + \tfrac{1}{2}\Sigma\frac{[\mathrm{tr}\,(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i)]^2}{\nu_i}\right]; \tag{8·25}$$

and $r = p(k-1)$ is the number of degrees of freedom of χ^2. Now

$$\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_1 = \begin{bmatrix} 0{\cdot}6585 & -0{\cdot}3028 \\ -0{\cdot}0970 & 0{\cdot}2751 \end{bmatrix}, \quad \mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_2 = \begin{bmatrix} 0{\cdot}6033 & 0{\cdot}2391 \\ 0{\cdot}0761 & 0{\cdot}8100 \end{bmatrix},$$

$$\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_3 = \begin{bmatrix} 0{\cdot}7382 & 0{\cdot}0637 \\ 0{\cdot}0209 & 0{\cdot}9149 \end{bmatrix}; \tag{8·26}$$

$$\mathrm{tr}\,(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_1)^2 = 0{\cdot}5680, \quad \mathrm{tr}\,(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_2)^2 = 1{\cdot}0565,$$
$$\mathrm{tr}\,(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_3)^2 = 1{\cdot}3846; \tag{8·27}$$

$$[\mathrm{tr}\,(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_1)]^2 = 0{\cdot}8716, \quad [\mathrm{tr}\,(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_2)]^2 = 1{\cdot}9974,$$
$$[\mathrm{tr}\,(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_3)]^2 = 2{\cdot}7327; \tag{8·28}$$

$$\Sigma\frac{\mathrm{tr}\,(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i)^2}{\nu_i} = 0{\cdot}2820, \quad \Sigma\frac{[\mathrm{tr}\,(\mathbf{I} - \mathbf{W}^{-1}\mathbf{W}_i)]^2}{\nu_i} = 0{\cdot}5311; \tag{8·29}$$

whence

$$A = 1 + \tfrac{1}{6}(0{\cdot}5311) = 1{\cdot}0664,$$
$$B = \tfrac{1}{24}[0{\cdot}2820 + \tfrac{1}{2}(0{\cdot}5311)] = 0{\cdot}02281. \tag{8·30}$$

The values of the multiplier $A + B\chi^2$ and of $2h(a)$ (to the order ν^{-1}) are given in Table 2. Our figure of 18·75 (equation (8·22)) is certainly *not* significant at the 0·1 % level; it may be so at the 1 % level, but such results should not of course be interpreted too literally, particularly when working far out in the tail of the distribution.

The author's thanks are again due to Mr A. M. Walker, who originally suggested, and first worked on, the problem of comparing two multivariate centroids; a discrepancy in our results led to the consideration of the multivariate linear hypothesis of § 5 as a means of checking this particular result. Mr Walker also checked the original draft of this paper, and pointed out a major error. However, the results have since been much extended, and the paper rewritten, so that any remaining mistakes are solely the author's responsibility.

REFERENCES

ASPIN, ALICE A. (1948). An examination and further development of a formula arising in the problem of comparing two mean values. *Biometrika*, **35**, 88–96.

ASPIN, ALICE A. (1949). Tables for use in comparisons whose accuracy involves two variances, separately estimated. *Biometrika*, **36**, 290–6.

JAMES, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika*, **38**, 324–9.

WELCH, B. L. (1947*a*). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, **34**, 28–35.

WELCH, B. L. (1947*b*). On the Studentization of several variances. *Ann. Math. Statist.* **18**, 118–22.

WELCH, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika*, **38**, 330–6.

WILKS, S. S. (1946). *Mathematical Statistics*. Princeton University Press.

THE USE OF THE HANKEL TRANSFORM IN STATISTICS

I. GENERAL THEORY AND EXAMPLES

By R. D. LORD

Royal Technical College, Glasgow

1. Introduction and summary

The use of the characteristic function in problems of the addition of independent random vectors is well established. This paper is concerned with the particular forms taken when the vectors have spherical distributions in s dimensions, i.e. when all directions are equally probable and the distribution of magnitudes is independent of direction. That the characteristic function is then a Hankel transform follows by changing to polar co-ordinates in the usual Cartesian form. A change to polars has sometimes been used in particular problems of this class but usually in a manner which does not show that the Hankel transform is the natural tool and that there is a theory for spherical distributions which runs parallel to the theory of the addition of random variables in one dimension. Soper's little known tract (Soper, 1922*) is an exception to this statement, for on pp. 38–47 he sketched a general procedure. However, Soper's work is rather inaccessible, very brief and uses a terminology different from that which has since become standard; he also uses the less convenient moment-generating function. A new account should therefore be useful, especially as the value of the Hankel transform in certain problems of mathematical physics is now known (Sneddon, 1951; Tranter, 1951).

The form of the characteristic function is obtained in §2. In §3 it is shown that it can also be obtained by the method used by Kluyver to solve Pearson's original random-walk problem. Another general procedure would be to use the appropriate form of the convolution. Although this method is not touched on here, it may be remarked that many of the integrals which arise in particular problems are evaluated in Watson (1944) by a disguised form of the method. The characteristic function is also a moment-generating function and §4 deals with this aspect. Since moments of order greater than two are not additive, cumulants are introduced as in the standard one-dimensional case.

Spherical distributions are often derived by projection from other spherical distributions in higher dimensional space. In §5 it is shown that distributions so related have the same characteristic function, and that solutions of s-dimensional problems can be derived quite simply from solutions in one or two dimensions. In §§6–9 are considered some special distributions and examples: normal, Cauchy and exponential. The random-walk problem of Pearson is left to another paper.

The special functions of mathematics have usually arisen first in potential problems, and arguments about potential are often used to provide physical pictures of analytical results. Our results can be used to provide similar pictures in terms of probability. Further, we

* I owe my acquaintance with it to Prof. Aitken, who told me of Soper's little book and kindly lent me his copy when I gave a paper at a meeting of the Edinburgh Mathematical Society on a solution of Pearson's problem using Hankel transforms (Lord, 1948). The main results of the present paper have been known to me since early in 1948; some of them were obtained independently by Dr N. B. Slater.

obtain integrals as answers to our problems in probability; reversing the viewpoint, a probability experiment may be a practical method of evaluating these integrals, i.e. the results given here suggest possible 'Monte Carlo' methods of evaluating integrals involving Bessel functions.

2. THE CHARACTERISTIC FUNCTION AND INVERSION THEOREMS

2·1. Let $\mathbf{X}$ be an s-dimensional random vector with probability function $p(\mathbf{X})$, i.e. let $p(\mathbf{X})\,d\mathbf{X}$ be the probability that the end-point of the vector lies in the volume element $d\mathbf{X}$ at the point whose position vector is $\mathbf{X}$. The characteristic function of the distribution of $\mathbf{X}$ is $\phi(\mathbf{T})$, the mean value of $\exp(i\mathbf{X}.\mathbf{T})$, where $\mathbf{T}$ is a variable but not random vector, and $\mathbf{X}.\mathbf{T}$ denotes the scalar product of $\mathbf{X}$ and $\mathbf{T}$, i.e.

$$\phi(\mathbf{T}) = \int \exp(i\mathbf{X}.\mathbf{T})\,p(\mathbf{X})\,d\mathbf{X}, \tag{1}$$

where the integration is over the whole space. This integral is absolutely convergent and this justifies its subsequent expression as a repeated integral.

Now let

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n,$$

where $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are independent random vectors with characteristic functions $\phi_1(\mathbf{T})$, $\phi_2(\mathbf{T}), \ldots, \phi_n(\mathbf{T})$. Then it is well known that

$$\phi(\mathbf{T}) = \phi_1(\mathbf{T})\,\phi_2(\mathbf{T}) \ldots \phi_n(\mathbf{T}). \tag{2}$$

This equation is fundamental in the process of finding the probability function of $\mathbf{X}$. To complete the process we use the inversion formula

$$p(\mathbf{X}) = (2\pi)^{-s} \int \exp(-i\mathbf{X}.\mathbf{T})\,\phi(\mathbf{T})\,d\mathbf{T}, \tag{3}$$

where now the integration is over the whole $\mathbf{T}$ space. We use these equations in the ordinary way, i.e. find ϕ_1, ϕ_2, etc., from (1), substitute in (2) to obtain ϕ and then in (3) to obtain $p(\mathbf{X})$. However, we compute (1) and (3) from analytical forms different from the usual Cartesian ones.

2·2. When $\mathbf{X}$ has a spherical distribution, $p(\mathbf{X})$ is a function of $|\mathbf{X}| = r$ only; write it as $p(r)$. The volume element is $d\mathbf{X} = dr\,dS$, where dS is an element of surface of the s-dimensional sphere of radius r, i.e.

$$d\mathbf{X} = r^{s-1}\,dr\,d\Sigma_s,$$

where Σ_s is the surface of the unit sphere in s-dimensions. Also

$$\mathbf{X}.\mathbf{T} = r\rho\cos\theta,$$

where $\rho = |\mathbf{T}|$ and θ is the angle between $\mathbf{X}$ and $\mathbf{T}$. Hence

$$\phi(\mathbf{T}) = \int_0^\infty r^{s-1} p(r)\,dr \int \exp(ir\rho\cos\theta)\,d\Sigma_s. \tag{4}$$

The inner integral is over Σ_s, and clearly its value is independent of the direction of $\mathbf{T}$ and is a function of $r\rho$ only. Thus $\phi(\mathbf{T})$ is a function of ρ only, say $\phi(\mathbf{T}) = \Phi(\rho)$.

To evaluate the inner integral we take polar co-ordinates $(r, \theta_1, \theta_2, \ldots, \theta_{s-1})$, where $\theta_1, \theta_2, \ldots, \theta_{s-2}$ run from 0 to π, and θ_{s-1} from 0 to 2π. Since the integral does not depend on

the direction **T** we can take **T** along $\theta_1 = 0$ and then $\theta = \theta_1$. In the important cases $s = 2, 3$, we obviously obtain (respectively)

$$\int_0^{2\pi} \exp(ir\rho\cos\theta_1)\,d\theta_1, \quad 2\pi\int_0^{\pi} \exp(ir\rho\cos\theta_1)\sin\theta_1\,d\theta_1,$$

i.e. $$2\pi J_0(r\rho), \quad (4\pi\sin r\rho)/(r\rho),$$

which substituted in (4) give the first equations of each of (31) and (32) below.

In the general case we use the transformation

$$\begin{aligned}
x_1 &= r\cos\theta_1,\\
x_2 &= r\sin\theta_1\cos\theta_2,\\
&\cdots \quad \cdots \quad \cdots\\
x_k &= r\sin\theta_1\sin\theta_2\ldots\sin\theta_{k-1}\cos\theta_k,\\
&\cdots \quad \cdots \quad \cdots\\
x_{s-1} &= r\sin\theta_1\sin\theta_2\ldots\sin\theta_{s-2}\cos\theta_{s-1},\\
x_s &= r\sin\theta_1\sin\theta_2\ldots\sin\theta_{s-2}\sin\theta_{s-1},
\end{aligned}$$

and easily find, using $$\Sigma_s = 2\pi^{\frac{1}{2}s}/\Gamma(\tfrac{1}{2}s),$$

that the inner integral in (4) has the value

$$(2\pi)^{\frac{1}{2}s}(r\rho)^{-\frac{1}{2}s+1}J_{\frac{1}{2}s-1}(r\rho).$$

We then have $$\Phi(\rho) = (2\pi)^{\frac{1}{2}s}\rho^{-\frac{1}{2}s+1}\int_0^{\infty} r^{\frac{1}{2}s}J_{\frac{1}{2}s-1}(r\rho)\,p(r)\,dr. \tag{5}$$

Since $\Phi(\rho)$ is a function of ρ only, the same process applied to (3) will give the inversion formula

$$p(r) = (2\pi)^{-\frac{1}{2}s}r^{-\frac{1}{2}s+1}\int_0^{\infty} \rho^{\frac{1}{2}s}J_{\frac{1}{2}s-1}(r\rho)\,\Phi(\rho)\,d\rho. \tag{6}$$

A different method of transformation will be found in Sneddon (1951) or in Bochner & Chandrasekharan (1949).

2·3. Instead of $p(r)$ we shall often use the function $P(r)$, where $P(r)\,dr$ is the probability that $|\mathbf{X}|$ lies between r and $r+dr$; $P(r)$ is thus the probability function of the distribution of $|\mathbf{X}|$, which is one-dimensional, while $p(r)$ refers to the s-dimensional distribution of $\mathbf{X}$. Clearly $$P(r) = r^{s-1}p(r)\,\Sigma_s = 2\pi^{\frac{1}{2}s}r^{s-1}p(r)/\Gamma(\tfrac{1}{2}s). \tag{7}$$

If we define $\Lambda_\alpha(z)$ with Jahnke & Emde (1945) and others, by

$$\Lambda_\alpha(z) = \Gamma(\alpha+1)\,(\tfrac{1}{2}z)^{-\alpha}J_\alpha(z) = 1 - \frac{1}{2(2\alpha+2)}z^2 + \frac{1}{2.4(2\alpha+2)(2\alpha+4)}z^4 - \ldots, \tag{8}$$

then (5) takes the very simple form

$$\Phi(\rho) = \int_0^{\infty} P(r)\,\Lambda_{\frac{1}{2}s-1}(r\rho)\,dr. \tag{9}$$

The inversion formula for $P(r)$ is

$$P(r) = 2^{-\frac{1}{2}s+1}\{\Gamma(\tfrac{1}{2}s)\}^{-1}\int_0^{\infty} (r\rho)^{\frac{1}{2}s}J_{\frac{1}{2}s-1}(r\rho)\,\Phi(\rho)\,d\rho. \tag{10}$$

Another alternative is to use the distribution function (or 'probability integral')

$$F(r) = \int_0^r P(u)\,du. \tag{11}$$

If we put $P(r) = F'(r)$ in (9) and integrate by parts, using $F(0) = 0$ and the fact that $\Lambda_{\frac{1}{2}s-1}(z)$ is of order $z^{-\frac{1}{2}s+\frac{1}{2}}$ at infinity, we obtain

$$\left.\begin{aligned}\Phi(\rho) &= (\rho^2/s)\int_0^\infty r\Lambda_{\frac{1}{2}s}(r\rho)\,F(r)\,dr \\ &= 2^{\frac{1}{2}s-2}\Gamma(\tfrac{1}{2}s)\,\rho^{2-\frac{1}{2}s}\int_0^\infty r^{1-\frac{1}{2}s}J_{\frac{1}{2}s}(r\rho)\,F(r)\,dr.\end{aligned}\right\} \tag{12}$$

The inverse formula is

$$F(r) = 2^{-\frac{1}{2}s+1}\{\Gamma(\tfrac{1}{2}s)\}^{-1}r^{\frac{1}{2}s}\int_0^\infty \rho^{\frac{1}{2}s-1}J_{\frac{1}{2}s}(r\rho)\,\Phi(\rho)\,d\rho. \tag{13}$$

2·4. Since $\Lambda_n(x)$ is a continuous function with unity for upper bound, attained when $x = 0$, it follows from (9) that $\Phi(\rho)$ is a continuous function of ρ and that

$$\Phi(0) = \int_0^\infty P(r)\,dr = 1, \tag{14}$$

$$|\,\Phi(\rho)\,| < 1 \quad (\rho > 0). \tag{15}$$

Also if (as is commonly the case, though the Cauchy distribution of §8 is an exception) we can integrate the right-hand side of (9) after expanding the Bessel function in a power series, $\Phi(\rho)$ will be an even function of ρ since $\Lambda_n(x)$ is even.

The analogue of the Riemann-Lebesgue lemma (Watson, 1944, p. 457) enables us to assert more than (15), namely,

$$\Phi(\rho) = o(\rho^{\frac{1}{2}-\frac{1}{2}s}),$$

as ρ tends to infinity, provided $\int_0^\infty r^{\frac{1}{2}-\frac{1}{2}s}P(r)\,dr$ exists; this integral converges at the upper limit but might fail to do so at the lower. For distributions with a $P(r)$ which is differentiable, the order of $\Phi(\rho)$ at infinity can be proved smaller by applying the Riemann-Lebesgue lemma after (repeated) integration by parts. Conversely, the order of $\Phi(\rho)$ at infinity provides us with information about the differentiability of $P(r)$.

3. Proof by the method of Kluyver and Markoff

3·1. We now show how the form of the characteristic function, its fundamental property (2) and the inversion formula could have been obtained by the method that Kluyver (1906) used to solve Pearson's problem; Kluyver's proof is reproduced with some generalization in Watson (1944). The corresponding proof for Cartesians was given by Chandrasekhar (1943), who then derived some special cases by a change of variables. Following his usage the method has often been called Markoff's method, but it was also used by Kluyver and may well be much older; there are indications that, following Kluyver, Heaviside used it in the missing fourth volume of his *Electromagnetic Theory*. It is inferior to the method of §2, at least in so far as it gives the final result in a form which obscures the importance of the characteristic function.

It is sufficient to deal with $\mathbf{X}$, the sum of two independent vectors $\mathbf{X}_1, \mathbf{X}_2$ with probability functions $p_1(\mathbf{X}), p_2(\mathbf{X})$. It will be shown that the distribution function $F(r)$ is given by (13) with $\Phi = \Phi_1\Phi_2$, where Φ is defined as in (5) and Φ_1, Φ_2 differ from Φ in having $p_1(r)$, $p_2(r)$ respectively instead of $p(r)$. This is equivalent to the results of §2.

The general idea of the method is as follows. To find the probability that $\mathbf{X}_1, \mathbf{X}_2$ satisfy a certain condition C we have to evaluate

$$\int p_1(\mathbf{X}_1)\, p_2(\mathbf{X}_2)\, d\mathbf{X}_1 d\mathbf{X}_2, \tag{16}$$

taken over all values of $\mathbf{X}_1, \mathbf{X}_2$ satisfying C. Let $\delta(\mathbf{X}_1, \mathbf{X}_2, C)$ be a function equal to unity when C is satisfied and zero otherwise. Then the required probability is

$$\int p_1(\mathbf{X}_1)\, p_2(\mathbf{X}_2)\, \delta(\mathbf{X}_1, \mathbf{X}_2, C)\, d\mathbf{X}_1, d\mathbf{X}_2,$$

where now the integration is over all values of $\mathbf{X}_1, \mathbf{X}_2$. If $\delta(\mathbf{X}_1, \mathbf{X}_2, C)$ can be expressed in the form of an integral, e.g. Markoff uses Dirichlet's discontinuous factor, (16) may be evaluable by changes of order of integration.

3·2. We require the probability that $|\mathbf{X}_1 + \mathbf{X}_2| < r$, i.e. that

$$R = (r_1^2 + r_2^2 - 2r_1 r_2 \cos\theta)^{\frac{1}{2}} < r,$$

where R, r_1, r_2 are the magnitudes of $\mathbf{X}_1 + \mathbf{X}_2, \mathbf{X}_1, \mathbf{X}_2$ respectively and θ is the angle between $\mathbf{X}_1$ and $\mathbf{X}_2$. Now

$$r^{\frac{1}{2}s} \int_0^\infty J_{\frac{1}{2}s}(r\rho)\, \frac{J_{\frac{1}{2}s-1}(R\rho)}{R^{\frac{1}{2}s-1}}\, d\rho = 1, \quad \text{if } R < r,$$
$$= 0, \quad \text{if } R > r.$$

This integral will be used as $\delta(\mathbf{X}_1, \mathbf{X}_2, C)$.

We first evaluate

$$I = \int p_2(\mathbf{X}_2)\, \delta(\mathbf{X}_1, \mathbf{X}_2, C)\, d\mathbf{X}_2$$

for a fixed value of $\mathbf{X}_1$. We use θ as a co-latitude in defining the position of $\mathbf{X}_2$, and put

$$p_2(\mathbf{X}_2)\, d\mathbf{X}_2 = p_2(r_2)\, r_2^{s-1} dr_2 d\Sigma_s$$
$$= p_2(r_2)\, r_2^{s-1} dr_2 \sin^{s-2}\theta\, d\theta\, d\Sigma_{s-1}.$$

We thus obtain

$$I = r^{\frac{1}{2}s} \int p_2(r_2)\, r_2^{s-1} J_{\frac{1}{2}s}(r\rho)\, \frac{J_{\frac{1}{2}s-1}(R\rho)}{R^{\frac{1}{2}s-1}} \sin^{s-2}\theta\, d\rho\, dr_2 d\theta\, d\Sigma_{s-1},$$

where the integration is over the surface of the unit sphere Σ_{s-1}, from 0 to π for θ and from 0 to ∞ for ρ and r_2.

Since the polar co-ordinates (other than r_2 and θ) needed to specify $\mathbf{X}_2$ enter only into $d\Sigma_{s-1}$, the integration over the sphere merely gives us a factor Σ_{s-1}. For the integration with respect to θ we use (Watson, 1944, p. 367 (16))

$$\int_0^\pi \frac{J_{\frac{1}{2}s-1}(R\rho)}{R^{\frac{1}{2}s-1}} \sin^{s-2}\theta\, d\theta = 2^{\frac{1}{2}s-1} \pi^{\frac{1}{2}} \Gamma(\tfrac{1}{2}s - \tfrac{1}{2})\, \rho^{1-\frac{1}{2}s}\, \frac{J_{\frac{1}{2}s-1}(\rho r_1)}{r_1^{\frac{1}{2}s-1}}\, \frac{J_{\frac{1}{2}s-1}(\rho r_2)}{r_2^{\frac{1}{2}s-1}}.$$

Substituting in I and carrying out the integration with respect to r_2,

$$I = r^{\frac{1}{2}s} \int_0^\infty \frac{J_{\frac{1}{2}s-1}(\rho r_1)}{r_1^{\frac{1}{2}s-1}}\, J_{\frac{1}{2}s}(r\rho)\, \Phi_2(\rho)\, d\rho.$$

Now substitute for I in (16) with $p_1(\mathbf{X}_1)\, d\mathbf{X}_1$ put equal to $p_1(r_1)\, r_1^{s-1} d\Sigma_s$:

$$F(r) = r^{\frac{1}{2}s} \int p_1(r_1)\, r_1^{s-1}\, \frac{J_{\frac{1}{2}s-1}(\rho r_1)}{r_1^{\frac{1}{2}s-1}}\, J_{\frac{1}{2}s}(r\rho)\, \Phi_2(\rho)\, d\rho\, dr_1 d\Sigma_s,$$

where the integration is over Σ_s and from 0 to ∞ for ρ and r_1. The integration over Σ_s gives a factor Σ_s and that with respect to r_1 gives

$$\int_0^\infty p_1(r_1)\, r_1^{\frac{1}{2}s} J_{\frac{1}{2}s-1}(\rho r_1)\, dr_1,$$

i.e. $$(2\pi)^{-\frac{1}{2}s}\rho^{\frac{1}{2}s-1}\Phi_1(\rho).$$

Thus $$F(r) = 2^{1-\frac{1}{2}s}\{\Gamma(\tfrac{1}{2}s)\}^{-1} r^{\frac{1}{2}s}\int_0^\infty \rho^{\frac{1}{2}s-1} J_{\frac{1}{2}s}(r\rho)\,\Phi_1(\rho)\,\Phi_2(\rho)\,d\rho,$$

and this is (13) with $\Phi = \Phi_1\Phi_2$.

4. MOMENTS AND CUMULANTS

4·1. The νth moment of the distribution of $\mathbf{X}$ is defined by the equation

$$\mu_\nu = \int_0^\infty P(r)\, r^\nu\, dr. \tag{17}$$

If in (9) we expand the Bessel function according to (8) and integrate term by term we obtain

$$\Phi(\rho) = 1 - \frac{1}{2s}\mu_2\rho^2 + \frac{1}{2.4s(s+2)}\mu_4\rho^4 - \ldots, \tag{18}$$

so that $\Phi(\rho)$ can be used as a moment-generating function. If only moments up to and including μ_{2n} exist then (18) will finish with a term $o(\mu_{2n}\rho^{2n})$. $\Phi(\rho)$ generates only even moments, but this is not surprising since we are dealing with symmetrical distributions. $P(r)$ has been defined to be unsymmetrical about $r = 0$, but it could be made symmetrical if we restricted the angular co-ordinate θ_{s-1} to run from 0 to π and allowed negative values of r.

4·2. Multiplying series such as (18) clearly shows that the second moment of the sum $(\mathbf{X}_1 + \mathbf{X}_2 + \ldots)$ is the sum of the second moments of $\mathbf{X}_1, \mathbf{X}_2, \ldots$, and that this additive property does not extend to the higher moments. The coefficients in the expansion of $\log\Phi(\rho)$ will, however, be additive. Hence we define *cumulants* $\lambda_2, \lambda_4, \lambda_6, \ldots$ by the equation

$$\log\Phi(\rho) = -\frac{1}{2s}\lambda_2\rho^2 + \frac{1}{2.4s(s+2)}\lambda_4\rho^4 - \frac{1}{2.4.6s(s+2)(s+4)}\lambda_6\rho^6 + \ldots, \tag{19}$$

$\log\Phi(\rho)$ being the cumulant-generating function. When $s = 1$, these reduce to the cumulants of a symmetrical one-dimensional distribution as usually defined (see, for example, Kendall, 1943). If in other cases we need to make a distinction we can use the term *polar cumulants*.

4.3. Formulae connecting moments and cumulants of lowest order are

$$\mu_2 = \lambda_2; \tag{20}$$

$$\left.\begin{aligned}
\mu_4 &= \lambda_4 + \frac{s+2}{s}\lambda_2^2,\\
\mu_6 &= \lambda_6 + 3\frac{s+4}{s}\lambda_4\lambda_2 + \frac{(s+2)(s+4)}{s^2}\lambda_2^3,\\
\mu_8 &= \lambda_8 + 4\frac{s+6}{s}\lambda_6\lambda_2 + 3\frac{(s+4)(s+6)}{s(s+2)}\lambda_4^2 + 6\frac{(s+4)(s+6)}{s^2}\lambda_4\lambda_2^2 + \frac{(s+2)(s+4)(s+6)}{s^3}\lambda_2^4,
\end{aligned}\right\} \tag{21}$$

$$\left.\begin{aligned}
\lambda_4 &= \mu_4 - \frac{s+2}{s}\mu_2^2, \\
\lambda_6 &= \mu_6 - 3\frac{s+4}{s}\mu_4\mu_2 + 2\frac{(s+2)(s+4)}{s^2}\mu_2^3, \\
\lambda_8 &= \mu_8 - 4\frac{s+6}{s}\mu_6\mu_2 - 3\frac{(s+4)(s+6)}{s(s+2)}\mu_4^2 + 12\frac{(s+4)(s+6)}{s^2}\mu_4\mu_2^2 - 6\frac{(s+2)(s+4)(s+6)}{s^3}\mu_2^4.
\end{aligned}\right\} \tag{22}$$

These and further formulae can be obtained by equating coefficients of powers of ρ in the equation obtained by substituting from (18) into (19). It is, however, easier to utilize the known results for the standard non-symmetrical case in one dimension. Write

$$\mu'_\nu = \frac{\mu_{2\nu}}{s(s+2)\dots(s+2\nu-2)}, \quad \kappa_\nu = \frac{\lambda_{2\nu}}{s(s+2)+(s+2\nu-2)}, \tag{23}$$

and also $t = -\frac{1}{2}\rho^2$. With this notation we are to equate coefficients in

$$\log\left(1+\mu'_1 t + \frac{1}{2!}\mu'_2 t^2 + \dots\right) = \kappa_1 t + \frac{1}{2!}\kappa_2 t^2 + \dots.$$

But this is precisely the equation that we use in the ordinary one-dimensional case (when μ'_ν is the νth moment about an arbitrary origin and κ_ν is the νth cumulant; see, for example, Kendall (1943), equation (3·22)). Consequently the formulae worked out for this standard case will give the formulae we seek if we replace μ'_ν, κ_ν by the expressions in (23). From equations (3·29) and (3·33) of Kendall (1943) may be obtained expressions for moments and cumulants up to and including μ_{20}, λ_{20}. It should be emphasized that the notation in (23) is purely temporary and intended only for the purposes of this paragraph.

5. Projected distributions

5·1. When a spherical distribution D_s in s dimensions is projected on to a subspace of q dimensions passing through the origin the result is obviously another spherical distribution D_q; for example, a three-dimensional distribution might be projected on to a diametral plane ($s = 3, q = 2$), or on to a diameter ($s = 3, q = 1$). Using Cartesian co-ordinates $x_1, x_2, \dots, x_q, \dots, x_s$ with $x_1, \dots, x_q$ as the co-ordinates in the space of the projection, the probability function for D_q is

$$p_q(x_1, x_2, \dots, x_q) = \int p_s(x_1, x_2, \dots, x_s)\, dx_{q+1} \dots dx_s, \tag{24}$$

with characteristic function

$$\begin{aligned}
\phi_q(t_1, t_2, \dots, t_q) &= \int p_q(x_1, x_2, \dots, x_q) \exp\left[i(t_1x_1 + t_2x_2 + \dots + t_qx_q)\right] dx_1 dx_2 \dots dx_q \\
&= \int p_s(x_1, x_2, \dots, x_s) \exp\left[i(t_1x_1 + t_2x_2 + \dots + t_qx_q)\right] dx_1 dx_2 \dots dx_s,
\end{aligned}$$

i.e.

$$\phi_q(t_1, t_2, \dots, t_q) = \phi_s(t_1, t_2, \dots, t_q, 0, \dots, 0), \tag{25}$$

where ϕ_s is the characteristic function of D_s. This holds for any distribution, spherical or not. However, if D_s is spherical, then both sides of (25) are functions of $\rho_1 = (t_1^2 + t_2^2 + \dots + t_q^2)^{\frac{1}{2}}$ only, say $\Phi_q(\rho_1)$ and $\Phi_s(\rho_1)$. Hence

$$\Phi_q(\rho_1) = \Phi_s(\rho_1), \tag{26}$$

i.e. if $\Phi(\rho)$ is the characteristic function of the spherical distribution D_s, then $\Phi(\rho_1)$ is the characteristic function of its projection D_q; in other words, a spherical distribution and its projections all have the same characteristic function.

5·2. From this there immediately follow relations between $\mu_2, \mu_4, \ldots$, the moments of D_s, and $m_2, m_4, \ldots$, the moments of the projection D_q. Expanding the two sides of (26) according to (18) and equating coefficients of powers of ρ_1,

$$\frac{\mu_2}{s} = \frac{m_2}{q}, \quad \frac{\mu_4}{s(s+2)} = \frac{m_4}{q(q+2)}, \quad \frac{\mu_6}{s(s+2)(s+4)} = \frac{m_6}{q(q+2)(q+4)}, \ldots, \tag{27}$$

with exactly similar relations between cumulants. For example, the distribution of a globular star-cluster is often spherical and has to be inferred from its projection on the 'plane of the sky'. The moments of the space distribution are obtainable from those of the plane distribution by applying (27) with $s = 3$, $q = 2$, and are $\frac{3}{2}m_2$, $\frac{15}{8}m_4$, $\frac{35}{16}m_6$, etc.

5·3. Equation (26) is one way of expressing the relation between D_s and its projection D_q. There are others in which the probability functions $p_s(r)$ and $p_q(r)$ are expressed in terms of one another. Change to polar co-ordinates in (24), letting r be radius vector in the space in which a point has co-ordinates $(x_1, x_2, \ldots, x_q)$, and u be radius vector in the space in which a point has co-ordinates $(x_{q+1}, x_{q+2}, \ldots, x_s)$. Then

$$p_q(r) = \int p_s(\sqrt{(r^2+u^2)})\, u^{s-q-1} d\Sigma_{s-q}\, du,$$

the integration being over the whole of the second space, i.e.

$$p_q(r) = \Sigma_{s-q} \int_0^\infty p_s(\sqrt{(r^2+u^2)})\, u^{s-q-1} du,$$

or putting $s = q+2m$ and $t^2 = r^2+u^2$,

$$p_q(r) = 2\pi^m \{\Gamma(m)\}^{-1} \int_r^\infty p_{q+2m}(t)\,(t^2-r^2)^{m-1}\, t\, dt. \tag{28}$$

To obtain $p_{q+2m}(r)$ in terms of $p_q(r)$, apply the operator $\left(-\frac{1}{r}\frac{d}{dr}\right)^m$ to (6) with $s = q$, using the recurrence relation

$$\left(-\frac{1}{z}\frac{d}{dz}\right)^m \frac{J_n(z)}{z^n} = \frac{J_{n+m}(z)}{z^{n+m}}.$$

The result is

$$p_{q+2m}(r) = \left(-\frac{1}{2\pi r}\frac{d}{dr}\right)^m p_q(r) \tag{29}$$

if m is an integer. If $2m = 1$, application of (28) and (29) in succession gives

$$p_{q+1}(r) = 2\int_r^\infty \frac{t p_{q+2}(t)\, dt}{\sqrt{(t^2-r^2)}} = -\frac{1}{\pi}\int_r^\infty \frac{p_q'(t)\, dt}{\sqrt{(t^2-r^2)}}, \tag{30}$$

in the proof of which we have regarded D_{q+1}, D_q as projections of the same D_{q+2}. Alternatively regarding D_{q+1}, D_q as having a common projection D_{q-1}

$$p_{q+1}(r) = -\frac{1}{2\pi r}\frac{dp_{q-1}}{dr} = -\frac{1}{\pi r}\frac{d}{dr}\left\{\int_0^\infty \frac{p_q(t)}{\sqrt{(t^2-r^2)}}\, dt\right\}, \tag{30a}$$

a form in which the integral is more likely to converge at the lower limit; divergence can occur only if $t = r$ is an infinity of $p_q'(t)$ in (30), or of $p_q(t)$ in (30a).

5·4. Equations (29), (30) and (30a) can be unified by the concept of integrals of any order. The appropriate definition of $I_\alpha f$, the integral of $f(x)$ of positive order α, is

$$I_\alpha f = \{\Gamma(\alpha)\}^{-1} \int_x^\infty (t-x)^{\alpha-1} f(t)\, dt.$$

It is well known (and easily verified) that

$$I_1 f = \int_x^\infty f(t)\, dt, \quad I_\alpha(I_\beta f) = I_{\alpha+\beta} f,$$

if the integrals are absolutely convergent, so that when α is an integer, $I_\alpha f$ is the α-fold integral of $f(x)$ as ordinarily defined. The operation inverse to I is differentiation with respect to $-x$ and we define a derivative of order α by

$$\frac{d^\alpha f}{d(-x)^\alpha} = \frac{d}{d(-x)} I_{1-\alpha} f \quad (0<\alpha<1),$$

$$= \frac{d^2}{d(-x)^2} I_{2-\alpha} f \quad (1<\alpha<2),$$

and so on.

If now we put $r^2 = R$, $t^2 = T$, $p_n(r) = f_n(R)$ in (28), we obtain

$$f_q(R) = \pi^m I_m f_{q+2m}(R). \tag{28a}$$

It then follows that

$$f_{q+2m}(R) = \pi^{-m} \frac{d^m}{d(-R)^m} f_q(R), \tag{29a}$$

which reduces to (29) when m is an integer, and to (30a) if $m = \frac{1}{2}$.

6. The cases $s = 1, 2, 3$

6·1. If $s = 1$, the vectors are confined to a straight line with a distribution symmetrical about the origin. If $p(x)$ is the probability function, then $P(r)$, the probability function for $r = |x|$, is equal to $2p(x)$ and the characteristic function is

$$\phi(t) = \int_{-\infty}^\infty p(x)\, e^{ixt}\, dx = \int_0^\infty P(r) \cos rt\, dr,$$

which agrees with (9), since $\Lambda_{-\frac{1}{2}}(z) = \cos z$.

If $s = 2$, the principal formulae are

$$\left.\begin{aligned} \Phi(\rho) &= 2\pi \int_0^\infty r J_0(r\rho)\, p(r)\, dr = \int_0^\infty J_0(r\rho)\, P(r)\, dr, \\ P(r) &= 2\pi r p(r) = \int_0^\infty r\rho J_0(r\rho)\, \Phi(\rho)\, d\rho, \\ F(r) &= r \int_0^\infty J_1(r\rho)\, \Phi(\rho)\, d\rho. \end{aligned}\right\} \tag{31}$$

If $s = 3$, the Bessel function involved is an elementary function, since $\Lambda_{\frac{1}{2}}(z) = \sin z/z$. The principal formulae are

$$\left.\begin{aligned} \Phi(\rho) &= (4\pi/\rho) \int_0^\infty r p(r) \sin r\rho\, dr = \int_0^\infty (\sin r\rho/r\rho)\, P(r)\, dr, \\ P(r) &= 4\pi r^2 p(r) = (2/\pi) \int_0^\infty r\rho \Phi(\rho) \sin r\rho\, d\rho, \\ F(r) &= (2r/\pi) \int_0^\infty \left(\frac{\sin r\rho}{r\rho} - \cos r\rho\right) \Phi(\rho)\, d\rho. \end{aligned}\right\} \tag{32}$$

7. THE NORMAL DISTRIBUTION

7·1. We take for our first example of a special distribution a case of the general normal distribution and briefly reconsider some of its properties in the light of our general theory.

A vector $\mathbf{X} = (x_1, x_2, \ldots, x_s)$ has a normal distribution if

$$p(\mathbf{X}) = C \exp\{-Q(x_1, x_2, \ldots, x_s)\},$$

where C is a constant and Q is a positive definite quadratic form in $x_1, x_2, \ldots, x_s$, which in the case of a spherical distribution must be a sum of squares with equal coefficients. Thus

$$p(r) = C \exp(-\tfrac{1}{2}r^2/A),$$

where A is a constant. Using Watson (1944, p. 394 (4)), we find for the characteristic function

$$\Phi(\rho) = C(2\pi A)^{\frac{1}{2}s} \exp(-\tfrac{1}{2}A\rho^2).$$

Then $\Phi(0) = 1$ gives $C = (2\pi A)^{-\frac{1}{2}s}$, and expanding in powers of ρ we find $A = \mu_2/s$, where μ_2 is the second moment of the distribution. Hence

$$p(r) = (s/(2\pi\mu_2))^{\frac{1}{2}s} \exp(-\tfrac{1}{2}sr^2/\mu_2), \tag{33}$$

$$P(r) = \frac{2}{\Gamma(\frac{1}{2}s)} \left(\frac{s}{2\mu_2}\right)^{\frac{1}{2}s} r^{s-1} \exp(-\tfrac{1}{2}sr^2/\mu_2), \tag{34}$$

$$\Phi(\rho) = \exp(-\tfrac{1}{2}\mu_2\rho^2/s). \tag{35}$$

Obviously all cumulants except $\lambda_2 = \mu_2$ are zero, and the higher moments are

$$\mu_{2\nu} = (s+2)(s+4)\ldots(s+2\nu-2)\, s^{1-\nu}\mu_2^{\nu}.$$

The distribution function is found by integrating (34) and when s is even is expressible in terms of exponential functions and when s is odd in terms of the 'normal probability integral'. Differences of this sort between odd and even s will be often found with other distributions.

7·2. By multiplication of characteristic functions it follows exactly as in the one-dimensional case that the sum of independent normally distributed random vectors is itself so distributed with second moment equal to the sum of second moments of the component vectors. Also since all spherical normal distributions have the same characteristic function (35), the difference being only in the scale factor μ_2/s, the projection of a spherical normal distribution is also normal.

The normal distribution has the same importance here as in the standard theory; it is the usual limiting form of the distribution of the sum of n independent vectors as n tends to infinity (central limit theorem). Taking the case when each vector has the same distribution with cumulants $\lambda_2 = \mu_2, \lambda_4, \ldots$, the sum will have cumulants $n\mu_2, n\lambda_4, \ldots$. Changing the unit so that the second moment becomes unity, the cumulants of the sum are $1, \lambda_4/(n\mu_2^2), \lambda_6/(n^2\mu_2^3), \ldots$, tending to limiting values $1, 0, 0, \ldots$, as n tends to infinity. These limiting values are the cumulants of a normal distribution. Thus the limiting probability function for large n will be given by

$$P(r) = \frac{2}{\Gamma(\frac{1}{2}s)} \left(\frac{s}{2n\mu_2}\right)^{\frac{1}{2}s} r^{s-1} \exp\left(\frac{-sr^2}{2n\mu_2}\right). \tag{36}$$

This equation gives an approximation which is often accurate enough for most purposes if n is as small as 10. It may be noted here that for distributions of the same class the approximation for a given n commonly improves as s increases.

The gaps in this deduction can probably be bridged by a more rigorous discussion using Hankel-Stieltjes transforms. We can certainly appeal to the known general theory, which shows that a sufficient condition for the validity of (36) is the existence of μ_3.

8. The Cauchy distribution

8·1. The Cauchy distribution in s dimensions with parameter a has probability function

$$p(r) = C(a^2+r^2)^{-\frac{1}{2}s-\frac{1}{2}} \quad (C = \text{constant}). \tag{37}$$

It answers the following question in geometrical probability: O is the foot of the perpendicular, of length a, from a given point A on to a given hyperplane (or line, if $s = 1$); lines are drawn at random through A to meet the hyperplane, all directions being equally likely; what is the probability that the line meets an element at distance r from O? If $s = 1$, the distribution is well enough known to figure in many text-books, but for $s > 1$ it may be new.

The present interest of the Cauchy distribution is that it provides an example for which the distribution of the sum of n vectors does not tend to the normal distribution. For using Watson (1944, p. 434 (2)) to evaluate the characteristic function and the constant C, we find that for all values of s

$$\Phi(\rho) = e^{-\rho a}. \tag{38}$$

Hence the sum of n vectors with Cauchy distributions with parameters $a_1, a_2, \ldots, a_n$ is a Cauchy distribution with parameter $(a_1+a_2+\ldots+a_n)$ and does not tend to normality as n tends to infinity. The result of §7·3 is inapplicable since no moments exist.

Also (38) shows immediately that the projection of a Cauchy distribution with parameter a is another Cauchy distribution with that parameter.

9. The exponential distribution

9·1. The distribution for which $p(r)$ is proportional to $\exp(-r/b)$ will be called an exponential distribution with parameter b. Using Watson (1944, p. 386 (6)) to evaluate the characteristic function, and (14) to evaluate the constant of proportionality, we find

$$p(r) = \pi^{\frac{1}{2}-\frac{1}{2}s}(2b)^{-s}e^{-r/b}/\Gamma(\tfrac{1}{2}s+\tfrac{1}{2}), \tag{39}$$

$$\Phi(\rho) = (1+b^2\rho^2)^{-\frac{1}{2}-\frac{1}{2}s}, \tag{40}$$

$$\mu_{2\nu} = s(s+1)\ldots(s+2\nu-1)\,b^{2\nu}, \tag{41}$$

$$\lambda_{2\nu} = 2^{\nu-1}(\nu-1)!\,(s+1)\,s(s+2)\ldots(s+2\nu-2)\,b^{2\nu}. \tag{42}$$

9·2. For the sum of n vectors, each with the same exponential distribution, the characteristic function is

$$\Phi_n(\rho) = (1+b^2\rho^2)^{-\frac{1}{2}n-\frac{1}{2}ns} \tag{43}$$

and inverting, using Watson (1944, p. 434 (2)), we obtain the probability function

$$p_n(r) = Ab^{-s}\left(\frac{r}{b}\right)^{\alpha}K_{\alpha}\left(\frac{r}{b}\right), \tag{44}$$

where

$$2\alpha = ns+n-s, \quad A = 2^{1-s-\alpha}\pi^{-\frac{1}{2}s}/\Gamma(\tfrac{1}{2}ns+\tfrac{1}{2}n),$$

and $K_{\alpha}(x)$ denotes the second kind of Bessel function of imaginary argument. Whenever α is half an odd integer, $p_n(r)$ reduces to $p(r)$ multiplied by a polynomial in r/b of degree $\alpha-\frac{1}{2}$ (Watson 1944, p. 80 (12)).

Comparing (40) and (43) we see that $\Phi_n(\rho)$ is the characteristic function of an exponential distribution in $(ns+n-1)$ dimensions. Hence the sum of n vectors with identical exponential distributions with parameter b is the projection of a vector in $(ns+n-1)$ dimensions also with an exponential distribution with parameter b.

When the vectors to be added have distributions with different parameters $b_1, b_2, \ldots$, the characteristic function of their sum will be

$$\Phi_n(\rho) = (1+b_1^2\rho^2)^{-\beta}(1+b_2^2\rho^2)^{-\beta}\ldots(1+b_n^2\rho^2)^{-\beta},$$

where $\beta = \frac{1}{2}s + \frac{1}{2}$. When s is even the integral for $p_n(r)$ seems intractable even for the simple case $s = 2 = n$. When s is odd, β is an integer and $\Phi_n(\rho)$ can be split into partial fractions each of which will contribute to $p_n(r)$ a term like that on the right-hand side of (44).

9·3. Distributions of type (44) with α real and positive provide a generalization of the exponential integral, which would appear to be the 'right' generalization to s dimensions of the gamma or Type III distribution of statistics. We may rewrite the probability function as

$$p(r) = Ab^{-s}(r/b)^{\alpha} K_{\alpha}(r/b), \tag{45}$$

with $A = 2^{1-s-\alpha}\pi^{-\frac{1}{2}s}/\Gamma(\alpha + \frac{1}{2}s)$, and the characteristic function as

$$\Phi(\rho) = (1 + b^2\rho^2)^{-\alpha-\frac{1}{2}s}. \tag{46}$$

The parameters α, b determine the distribution, b being a scale parameter; $\alpha = \frac{1}{2}$ gives the exponential distribution.

By multiplication of characteristic functions it is clear that the sum of two vectors whose distributions have parameters α_1, b and α_2, b has a distribution of the same type with parameters $\alpha_1 + \alpha_2 + \frac{1}{2}s, b$. It is also clear from (46) that distributions with the same b and the same $(\alpha + \frac{1}{2}s)$ are related by projection.

The distribution function for (45) is

$$F(r) = B\int_0^{r/b} u^{\alpha+s-1} K_{\alpha}(u)\, du, \tag{47}$$

where $B = 2^{2-s-\alpha}/\{\Gamma(\frac{1}{2}s)\,\Gamma(\alpha + \frac{1}{2}s)\}$. If $s = 2$, this is

$$F(r) = 1 - B(r/b)^{\alpha+1} K_{\alpha+1}(r/b), \tag{48}$$

where $B = 2^{-\alpha}/\Gamma(\alpha + 1)$, since

$$\frac{d}{du}\{u^n K_n(u)\} = -u^n K_{n-1}(u).$$

If s is any other even number, $F(r)$ can be similarly obtained by integration by parts. If s is odd, the integration can be made to depend on the evaluation of

$$\int_0^x u^{\beta} K_{\beta}(u)\, du,$$

which has been tabulated by Pearson and others for $\beta = 0(\frac{1}{2})11$ (see Fletcher, Miller & Rosenhead (1946) for precise references); Pearson has also tabulated $u^{\beta}K_{\beta}(u)$ for the same range of values of β. If α is half an odd integer we can evaluate $1 - F(r)$ for all values of s as $e^{-r/b}$ multiplied by a polynomial of degree $\alpha + s - \frac{3}{2}$.

REFERENCES

BOCHNER, S. & CHANDRASEKHARAN, K. (1949). *Fourier Transforms*. Annals of Math. Studies, no. 19. Princeton.

CHANDRASEKHAR, S. (1943). Stochastic problems in physics and astronomy. *Rev. Mod. Phys.* **15**, 1–89.

FLETCHER, A., MILLER, J. C. P. & ROSENHEAD, L. (1946). *An Index of Mathematical tables*. London.

JAHNKE, E. & EMDE, F. (1945). *Tables of Functions with Formulae and Curves*, 4th ed. New York.

KENDALL, M. G. (1943). *The Advanced Theory of Statistics*, **1**. London.

KLUYVER, J. C. (1906). A local probability problem. *Proc. Section of Sci., K. Akad. Wet. Amst.* **8**, 341–50.

LORD, R. D. (1948). A problem on random vectors. *Phil. Mag.* (7), **39**, 66–71.

SNEDDON, I. N. (1951). *Fourier Transforms*. New York.

SOPER, H. E. (1922). *Frequency Arrays*. Ashford, Kent.

TRANTER, C. J. (1951). *Integral Transforms in Mathematical Physics*. London.

WATSON, G. N. (1944, 1st ed. 1922). *A Treatise on the Theory of Bessel Functions*, 2nd ed. Cambridge University Press.

A NOTE ON THE CONSISTENCY AND MAXIMA OF THE ROOTS OF LIKELIHOOD EQUATIONS

By K. C. CHANDA, *Bombay University*

1. Introduction

An exhaustive study of the properties of the maximum-likelihood estimates was made earlier by Huzurbazar (1948). He did, however, confine his discussions to distributions with only one unknown parameter. Although the extension of his results to multi-parametric distributions is conceptually easy, the algebraic details of such derivations are sometimes complex and at times quite interesting to follow. The consistency of the maximum-likelihood estimates for the general case was, however, deduced later by Wald (1949), but on the assumption that the estimate really maximizes the likelihood function; it is thus different from the maximum-likelihood estimate as defined by Fisher. No such assumptions are made here; but the basic assumptions underlying the proofs are completely different from, and in a sense stronger than, those of Wald.

2. Assumptions

Suppose $f(x, \theta)$ be the probability density law; $\theta = (\theta_1, \ldots, \theta_k)$ is the unknown parameter vector and let $x_1, \ldots, x_n$ be n independent observations on x. The likelihood equations for estimating θ are then given by

$$\frac{\partial \log \phi}{\partial \theta_r} = 0 \quad (r = 1, 2, \ldots, k)$$

where

$$\log \phi = \sum_{i=1}^{n} \log f(x_i, \theta).$$

For brevity's sake we shall henceforth write f for $f(x, \theta)$ and f_i for $f(x_i, \theta)$.

Let us now assume that:

(i) The point represented by the vector θ lies in a k-dimensional interval Ω; for almost all x and for all $\theta \in \Omega$

$$\frac{\partial \log f}{\partial \theta_r}, \quad \frac{\partial^2 \log f}{\partial \theta_r \partial \theta_s} \quad \text{and} \quad \frac{\partial^3 \log f}{\partial \theta_r \partial \theta_s \partial \theta_t}$$

exist for all $r, s, t = 1, 2, \ldots, k$.

(ii) For almost all x and for every point $\theta \in \Omega$

$$\left|\frac{\partial f}{\partial \theta_r}\right| < F_r(x), \quad \left|\frac{\partial^2 f}{\partial \theta_r \partial \theta_s}\right| < F_{rs}(x) \quad \text{and} \quad \left|\frac{\partial^3 \log f}{\partial \theta_r \partial \theta_s \partial \theta_t}\right| < H_{rst}(x),$$

while

$$\int_{-\infty}^{\infty} H_{rst}(x) f \, dx < M$$

for all $\theta \in \Omega$ and for all $r, s, t = 1, 2, \ldots, k$, M being a finite positive constant.

(iii) For all $\theta \in \Omega$ the matrix $J = ((J_{rs}(\theta)))$, where

$$J_{rs}(\theta) = \int_{-\infty}^{\infty} \frac{\partial \log f}{\partial \theta_r} \frac{\partial \log f}{\partial \theta_s} f \, dx$$

is positive-definite and that $|J|$ is finite.

Assumption (iii) forms the pivot of later proofs; but it is not, as it stands, very much restrictive. The reason is that, whenever we are studying the properties of different estimates (which include maximum-likelihood estimates as well), we have always at the background the concept of efficiency which in turn is measured by a simple function of the information-determinant $|J|$. For the class of unbiased estimates, the generalized variance is bounded below by a quantity which assumes the trivial value, zero, when $|J| = \infty$. This assumption is thus felt to be of quite general importance. As J is a dispersion matrix it is always positive-definite, except perhaps in degenerate cases. This may be proved very simply as follows:

Let $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ be any real row vector. Clearly

$$E\left\{\left[\lambda_1\frac{\partial \log f}{\partial \theta_1}+\lambda_2\frac{\partial \log f}{\partial \theta_2}+\ldots+\lambda_k\frac{\partial \log f}{\partial \theta_k}\right]\right\}^2 \geqslant 0,$$

i.e.

$$\Sigma\Sigma\lambda_r\lambda_s J_{rs} \geqslant 0,$$

which proves $[J]$ to be positive-definite.

3. Derivation of results

Let θ^0 be the unknown true value of the parameter vector θ, where we suppose that θ^0 is an inner point of Ω. Consider the following expansion:

$$\frac{\partial \log f}{\partial \theta_r} = \left(\frac{\partial \log f}{\partial \theta_r}\right)_{\theta=\theta^0} + \sum_{s=1}^{k}(\theta_s-\theta_s^0)\left(\frac{\partial^2 \log f}{\partial \theta_r \partial \theta_s}\right)_{\theta=\theta^0} + \frac{1}{2}\sum_{s,t=1}^{k}(\theta_s-\theta_s^0)(\theta_t-\theta_t^0)\left(\frac{\partial^3 \log f}{\partial \theta_r \partial \theta_s \partial \theta_t}\right)_{\theta=\theta'}, \quad (3{\cdot}1)$$

where $\theta' = \theta'(x)$ is a value depending on x, but for all x, lying inside the hyper-cell, of which the vector $\theta - \theta^0$ is the diagonal. Multiplying both sides by $1/n$ and summing the corresponding expressions for x_i's over $i = 1, 2, \ldots, n$ we may rewrite (3·1) as

$$L_r(\theta) = L_r(\theta^0) - \sum_{s=1}^{k}\delta_s L_{rs}(\theta^0) + \frac{1}{2}\sum_{s,t=1}^{k}\delta_s\delta_t L_{rst} \quad (r = 1, 2, \ldots, k), \quad (3{\cdot}2)$$

where

$$\left.\begin{aligned} \delta_s &= (\theta_s - \theta_s^0), \\ L_r(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial \log f_i}{\partial \theta_r}\right) = \frac{1}{n}\frac{\partial \log \phi}{\partial \theta_r}, \\ L_{rs}(\theta) &= -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2 \log f_i}{\partial \theta_r \partial \theta_s}\right) = -\frac{1}{n}\frac{\partial^2 \log \phi}{\partial \theta_r \partial \theta_s}, \\ L_{rst} &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^3 \log f_i}{\partial \theta_r \partial \theta_s \partial \theta_t}\right)_{\theta=\theta'(x_i)} \end{aligned}\right\} \quad (3{\cdot}3)$$

We note that

$$\left.\begin{aligned} &\text{(i)} \quad E\left(\frac{\partial \log f}{\partial \theta_r}\right) = 0 && (r = 1, 2, \ldots, k), \\ &\text{(ii)} \quad E\left(-\frac{\partial^2 \log f}{\partial \theta_r \partial \theta_s}\right) = J_{rs}(\theta) && (r, s = 1, 2, \ldots, k), \\ &\text{(iii)} \quad E\left(\left|\frac{\partial^3 \log f}{\partial \theta_r \partial \theta_s \partial \theta_t}\right|\right) < M && (r, s, t = 1, 2, \ldots, k), \end{aligned}\right\} \quad (3{\cdot}4)$$

for all $\theta \in \Omega$.

From (iii) it follows immediately that

$$E\left|\frac{\partial^3 \log f}{\partial\theta_r \partial\theta_s \partial\theta_t}\right|_{\theta=\theta'(x)} < M \quad \text{for all } x$$

(since $\theta'(x) \in \Omega$ for all x). Also

$$|L_{rst}| \leqslant \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\partial^3 \log fi}{\partial\theta_r \partial\theta_s \partial\theta_t}\right|_{\theta=\theta'(x_i)} < \frac{1}{n}\sum_{i=1}^{n} H_{rst}(x_i)$$

and

$$E\{H_{rst}(x_i)\} < M$$

for all $i = 1, 2, \ldots, n$; so by Khintchine's theorem,

$$\left.\begin{aligned} &\text{(i)}\ \ L_r(\theta^0) \to 0, \\ &\text{(ii)}\ \ L_{rs}(\theta^0) \to J_{rs}(\theta^0), \\ &\text{(iii)}\ \ |L_{rst}| < \frac{1}{n}\sum_{i=1}^{n} H_{rst}(x_i) \to E\{H_{rst}(x_i)\} < M, \quad \text{i.e. } |L_{rst}| \to B < M, \end{aligned}\right\} \tag{3·5}$$

where B is a finite-positive constant, with probability tending to unity. That is to say, we can choose an $n_0 = n_0(\eta, \epsilon)$, such that for all $n > n_0(\eta, \epsilon)$

$$\Pr[|L_r(\theta^0)| < \eta;\ |L_{rs}(\theta^0) - J_{rs}(\theta^0)| < \eta;\ |L_{rst}| < N] > 1 - \epsilon, \tag{3·6}$$

N being a finite-positive quantity greater than M and η and ϵ being two arbitrarily chosen small positive quantities. The likelihood equations are therefore given by putting $L_r(\theta) = 0$ $(r = 1, 2, \ldots, k)$ in (3·2), i.e.

$$L_r(\theta^0) - \sum_{s=1}^{k} \delta_s L_{rs}(\theta^0) + \frac{1}{2}\sum_{s,t=1}^{k} \delta_s \delta_t L_{rst} = 0 \quad (r = 1, 2, \ldots, k), \tag{3·7}$$

i.e.

$$\sum_{s=1}^{k} \delta_s L_{rs}(\theta^0) = L_r(\theta^0) + \frac{1}{2}\sum_{s,t=1}^{k} \delta_s \delta_t L_{rst} \quad (r = 1, 2, \ldots, k).$$

Again, since $E(L_{rs}(\theta^0)) = J_{rs}(\theta^0)$ and since $((J_{rs}(\theta^0)))^{-1} = J_0^{-1}$ (say) exists, $((L_{rs}(\theta^0)))^{-1} = L^{-1}$ (say) exists, too, and so we have from (3·7)

$$\delta_r = \beta_r + \sum_{s,t=1}^{k} \delta_s \delta_t a_{rst}, \tag{3·8}$$

where

$$\beta_r = \sum_{p=1}^{k} L_p(\theta^0) L^{pr}(\theta^0),$$

$$a_{rst} = \sum_{p=1}^{k} L_{pst} L^{pr}(\theta^0),$$

$$((L^{pr}(\theta^0))) = L^{-1}.$$

Since L^{-1} is a matrix with finite elements, we can choose our $n_0(\eta, \epsilon)$ for arbitrary η, ϵ such that for all $n > n_0(\eta, \epsilon)$

$$\Pr\left[|\beta_r| < \sum_{p=1}^{k} |L_p(\theta^0)|\,|L^{pr}(\theta^0)| < \eta;\right.$$
$$\left.|a_{rst}| < \frac{1}{2}\sum_{p=1}^{k} |L_{pst}|\,|L^{pr}(\theta^0)| < T \quad \text{for all } r = 1, 2, \ldots, k\right]$$
$$> 1 - \epsilon \quad \text{for} \quad \infty > T > N. \tag{3·9}$$

Consider now the set of equations

$$\delta_r = \beta_r + \sum_{s,t=1}^{k} \delta_s \delta_t a_{rst} \quad (r = 1, 2, \ldots, k).$$

It is easily seen that this set of equations admits a solution $\hat{\delta} = (\hat{\delta}_1, \ldots, \hat{\delta}_k)$ which are of the same order as η. For if we take η to be arbitrarily small, then in view of (3·9) it follows that the contribution due to the second term on the right-hand side can be made to be equal to a quantity of smaller order than η; and so

$$\Pr[|\hat{\delta}| < \eta] > 1 - \epsilon$$

for a sufficiently large value of n and for all values larger than that. It follows immediately, therefore, that there exists at least one solution of the likelihood equations, which is a consistent estimate of the true parameter vector θ^0. It is not, however, at once evident that this equation does not admit any other solution which may be inconsistent. We propose, however, to show that the consistent solution is unique. But before that we shall prove the following

THEOREM 1. The matrix $\Theta = \left(\left(\frac{\partial^2 \log \phi}{\partial\theta_r \partial\theta_s}\right)_{\theta=\hat{\theta}}\right)$, where $\hat{\theta}$ is a consistent root of the likelihood equation, is negative-definite with probability tending to unity.

Proof. Consider, in fact, the relation

$$\left(\frac{\partial^2 \log \phi}{\partial\theta_r \partial\theta_s}\right)_{\theta=\hat{\theta}} = \left(\frac{\partial^2 \log \phi}{\partial\theta_r \partial\theta_s}\right)_{\theta=\theta^0} + \sum_{t=1}^{k} (\hat{\theta}_t - \theta_t^0) \left(\frac{\partial^3 \log \phi}{\partial\theta_r \partial\theta_s \partial\theta_t}\right)_{\theta=\theta''}, \tag{3·10}$$

where as before $\theta'' = \theta''(x_1, \ldots, x_n)$ is, for all $(x_1, \ldots, x_n)$ an inner point of the hyper-cell of which the vector $\hat{\theta} - \theta^0$ is the diagonal.

We have

$$\left|\frac{1}{n}\left(\frac{\partial^2 \log \phi}{\partial\theta_r \partial\theta_s}\right)_{\theta=\hat{\theta}} - \frac{1}{n}\left(\frac{\partial^2 \log \phi}{\partial\theta_r \partial\theta_s}\right)_{\theta=\theta^0}\right| \leqslant \sum_{t=1}^{k} |\hat{\theta}_t - \theta_t^0| \left|\frac{1}{n}\left(\frac{\partial^3 \log \phi}{\partial\theta_r \partial\theta_s \partial\theta_t}\right)_{\theta=\theta''}\right|. \tag{3·11}$$

Since $\hat{\theta}$ is a consistent estimate of θ^0, and since $\theta''(x_1, \ldots, x_n)$ is an interior point of Ω,

$$\left.\begin{aligned} \Pr\Big[|\hat{\theta}_t - \theta_t^0| < \eta/2kN \ \ (t = 1, 2, \ldots, k); \\ \frac{1}{n}\left|\frac{\partial^3 \log \phi}{\partial\theta_r \partial\theta_s \partial\theta_t}\right|_{\theta=\theta''} < N\Big] > (1-\epsilon) \quad \text{for all } n > n_1(\eta, \epsilon). \end{aligned}\right\} \tag{3·12}$$

Hence for $n > n_1(\eta, \epsilon)$, we have

$$\Pr\left[\left|\frac{1}{n}\left(\frac{\partial^2 \log \phi}{\partial\theta_r \partial\theta_s}\right)_{\theta=\hat{\theta}} - \frac{1}{n}\left(\frac{\partial^2 \log \phi}{\partial\theta_r \partial\theta_s}\right)_{\theta=\theta^0}\right| < \tfrac{1}{2}\eta\right] > 1 - \epsilon. \tag{3·13}$$

Also $E\left(\frac{\partial^2 \log f}{\partial\theta_r \partial\theta_s}\right)_{\theta=\theta^0} = -J_{rs}(\theta^0)$, and so, by Khintchine's theorem, it follows that

$$\Pr\left[\left|\frac{1}{n}\left(\frac{\partial^2 \log \phi}{\partial\theta_r \partial\theta_s}\right)_{\theta=\theta^0} + J_{rs}(\theta^0)\right| < \tfrac{1}{2}\eta\right] > 1 - \epsilon \tag{3·14}$$

for $n > n_2(\eta, \epsilon)$.

Again we have

$$\left|\frac{1}{n}\left(\frac{\partial^2 \log \phi}{\partial\theta_r \partial\theta_s}\right)_{\theta=\hat{\theta}} + J_{rs}(\theta^0)\right| \leqslant \left|\frac{1}{n}\left(\frac{\partial^2 \log \phi}{\partial\theta_r \partial\theta_s}\right)_{\theta=\hat{\theta}} - \frac{1}{n}\left(\frac{\partial^2 \log \phi}{\partial\theta_r \partial\theta_s}\right)_{\theta=\theta^0}\right| + \left|\frac{1}{n}\left(\frac{\partial^2 \log \phi}{\partial\theta_r \partial\theta_s}\right)_{\theta=\theta^0} + J_{rs}(\theta^0)\right|.$$

Hence for $n > n_3 = \max(n_1, n_2)$, we have

$$\Pr\left[\left|\frac{1}{n}\left(\frac{\partial^2 \log \phi}{\partial\theta_r \partial\theta_s}\right)_{\theta=\hat{\theta}} + J_{rs}(\theta^0)\right| < \tfrac{1}{2}\eta + \tfrac{1}{2}\eta = \eta\right] > 1-\epsilon. \tag{3·15}$$

Consider the positive-definite quadratic form

$$Q_0 = \sum_{r,s=1}^{k} u_r u_s J_{rs}(\theta^0).$$

Let further

$$Q = \sum_{r,s=1}^{k} u_r u_s \Theta_{rs},$$

where

$$\Theta_{rs} = \frac{1}{n}\left(\frac{\partial^2 \log \phi}{\partial\theta_r \partial\theta_s}\right)_{\theta=\hat{\theta}}.$$

We have for $n > n_3$,

$$\Pr\left[-Q_0 - \eta(\Sigma u_r)^2 < Q < -Q_0 + \eta(\Sigma u_r)^2\right] > 1-\epsilon. \tag{3·16}$$

Since

$$Q_0 = \sum_{r,s=1}^{k} u_r u_s J_{rs}(\theta^0) \quad \text{and} \quad Q_0' = \left(\sum_{r=1}^{k} u_r\right)^2$$

are respectively of rank k and 1, it follows that we can by a non-singular linear transformation $u_r \to U_r$ convert Q_0 and Q_0' to $\sum_{r=1}^{k} U_r^2$ and μU_1^2 respectively, where $\mu > 0$ is the only non-zero root of the characteristic equation $|A - \mu J_0| = 0$, A being the matrix with all unit elements. Evidently $0 < \mu < \infty$, and so

$$\frac{Q_0}{\left(\sum_{r=1}^{k} u_r\right)^2} = \frac{1}{\mu} + \frac{\sum_{r=2}^{k} U_r^2}{\mu U_1^2} > \frac{1}{\mu} > \xi \text{ (say)}. \tag{3·17}$$

By taking $\eta = \xi$ and by choosing an $n_3(\eta, \epsilon)$ we can ensure that

$$\Pr[Q \leqslant 0] > 1-\epsilon \quad \text{for all } n > n_3(\eta, \epsilon)$$

and so

$$\Pr[\Theta \text{ is negative-definite}] > 1-\epsilon \quad \text{for all } n > n_3(\eta, \epsilon). \tag{3·18}$$

It follows immediately that the likelihood function has a relative maximum at $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ with probability tending to unity. Now we prove the following

THEOREM 2. Of all possible solutions to equations (3·8), one and only one tends in probability to the true parameter vector θ^0.

Proof. Suppose, if possible, $\hat{\theta}'$ be another consistent estimate and at the same time a solution of (3·8). By hypothesis,

$$\left(\frac{\partial \log \phi}{\partial\theta_r}\right)_{\theta=\hat{\theta}} = 0, \quad \left(\frac{\partial \log \phi}{\partial\theta_r}\right)_{\theta=\hat{\theta}'} = 0 \quad \text{for } r = 1, 2, \ldots, k. \tag{3·19}$$

Now let us consider only one such pair of equations, say the rth. It follows, therefore, in virtue of Rolle's theorem extended to multivariate functions that there exists a point $\hat{\theta}_r''$ (depending on r), lying inside the hyper-cell of which the vector $\hat{\theta} - \hat{\theta}'$ is the diagonal, such that

$$\left(\frac{\partial^2 \log \phi}{\partial\theta_r \partial\theta_s}\right)_{\theta=\hat{\theta}_r''} = 0 \quad \text{for } s = 1, 2, \ldots, k. \tag{3·20}$$

It is therefore immediately evident that the determinant of the matrix $\left(\left(\frac{\partial^2 \log \phi}{\partial\theta_p \partial\theta_s}\right)_{\theta=\hat{\theta}''_r}\right)$ vanishes. This, however, is a contradiction, since $\hat{\theta}''_r$ being a consistent estimate of θ^0 (since it lies inside the hyper-cell with diagonal $\hat{\theta}-\hat{\theta}'$ and since $\hat{\theta}$, $\hat{\theta}'$ are both consistent estimates of θ^0), the Hermitean $\left(\left(\frac{\partial^2 \log \phi}{\partial\theta_p \partial\theta_s}\right)_{\theta=\hat{\theta}''_r}\right)$ must be negative-definite with probability tending to unity, as we have seen earlier; hence the theorem is proved.

Consider now the likelihood equations

$$\sum_{s=1}^{k} \delta_s L_{rs}(\theta^0) = L_r(\theta^0) + \frac{1}{2} \sum_{s,t=1}^{k} \delta_s \delta_t L_{rst} \quad (r = 1, 2, \ldots, k). \tag{3·21}$$

We remember that:

(1) $L_r(\theta^0) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial \log f_i}{\partial\theta_r}\right)_{\theta=\theta^0}$, being the average of n independent variables, each having identical distribution, is itself distributed asymptotically normally with zero mean and variance $= J_{rr}(\theta^0)/n$. Again in virtue of the generalization of Liapounoff's central limit theorem, it follows that $L_1(\theta^0), \ldots, L_k(\theta^0)$ are asymptotically jointly distributed as a k-variate normal distribution, with zero means and variance-covariance matrix J_0/n.

(2) $L_{rs}(\theta^0) \to J_{rs}(\theta^0)$ with probability tending to unity.

(3) The second expression on the right-hand side of (3·21) is of a smaller order than $\sum_{s=1}^{k} \delta_s L_{rs}(\theta^0)$.

It follows, therefore, in virtue of Khintchine's theorem, that δ_r's have asymptotically a joint k-variate normal distribution with zero means and the variance-covariance matrix given by $V = ((nJ_{rs}(\theta^0)))^{-1}$.

REFERENCES

HUZURBAZAR, V. S. (1948). The likelihood equation, consistency, and the maxima of the likelihood function. *Ann. Eugen., Lond.*, **14**, 185.

WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20**, 595.

GROUPING METHODS IN THE FITTING OF POLYNOMIALS TO EQUALLY SPACED OBSERVATIONS

By P. G. GUEST
University of Sydney, Australia

Introduction

The fitting of polynomials to equally spaced observations by the method of least squares is carried out by means of power moments (Guest, 1953) or orthogonal moments (Anderson & Houseman, 1942), tables of orthogonal polynomials and related functions being used. When the number of observations is large (greater than 104) such tables are no longer available, and it becomes necessary to group the observations before performing the calculations. Even if the number of observations is less than 104 it may be considered advisable to group the observations in order to reduce the time spent on the calculations. It is the purpose of Part I of the present paper to show how unbiased estimates of the polynomial coefficients and fitted values may be obtained by least-squares methods when the observations are grouped, and to determine the reduction in efficiency which occurs.

A method of grouping which is of some interest is that in which step functions are used to evaluate the polynomial coefficients. The practical procedure using single-step functions has been outlined in an earlier paper (Guest, 1951). In Part II of the present paper the method of determining the optimum step functions will be discussed for both single-step and double-step functions. The fitting of a straight line by the use of step functions, being the most important case in practice, will be treated in detail.

Part I. Least-squares methods

(1) *The orthogonal coefficients*

The n observations $y(\epsilon)$ are first converted into N groups each containing r observations and represented by the symbols $Y(E)$. The values of the independent variables ϵ and E corresponding to the observations are supposed to be spaced at unit intervals in the range $+\frac{1}{2}(n-1)$ to $-\frac{1}{2}(n-1)$ and the range $+\frac{1}{2}(N-1)$ to $-\frac{1}{2}(N-1)$ respectively. Then

$$Y(E) = \sum_{z=-\frac{1}{2}(r-1)}^{+\frac{1}{2}(r-1)} y(rE+z). \tag{1}$$

A least-squares polynomial is then fitted to the N grouped values $Y(E)$, using either orthogonal moments or power moments. The fitted polynomial is written

$$U_p(E) = \sum_{j=0}^{p} A_j T_{jN}(E), \tag{2}$$

where $T_{jN}(E)$ (often written $\xi_{jN}(E)$) is the orthogonal polynomial of degree j in E with leading coefficient unity. The coefficients A_j are given by

$$A_j = \sum_E T_{jN}(E)\, Y(E) / \sum_E T_{jN}^2(E). \tag{3}$$

The problem is to obtain from these values A_j estimates a_j of the coefficients in the polynomial

$$u_p(\epsilon) = \sum_{j=0}^{p} a_j T_{jn}(\epsilon), \tag{4}$$

which is to fit the n original observations $y(\epsilon)$.

It is easy to show (Guest, 1952, p. 239) that if any arbitrary independent set of $p+1$ functions $W_j(\epsilon)$ is chosen the system of equations

$$\sum_{\epsilon} W_j(\epsilon)\,[y(\epsilon) - \sum_k a_k T_{kn}(\epsilon)] = 0 \tag{5}$$

leads to unbiased estimates of the polynomial coefficients. Since it is desired to express the a_k in terms of the A_k given by equation (3), the functions appropriate to the present discussion are

$$W_j(\epsilon) = T_{jN}(E), \quad rE - \frac{r-1}{2} \leqslant \epsilon \leqslant rE + \frac{r-1}{2},$$

and the system of equations determining a_k becomes

$$\sum_{E} T_{jN}(E)\,[Y(E) - \sum_k a_k \sum_z T_{kn}(rE+z)] = 0. \tag{6}$$

To simplify this expression it is necessary to expand $\sum_z T_{kn}(rE+z)$ in terms of the polynomials $T_{jN}(E)$. Taking as an example the case $k = 3$,

$$\begin{aligned}
\sum_z T_{3n}(rE+z) &= \sum_z [(rE+z)^3 - \{(3n^2-7)/20\}\,(rE+z)] \\
&= \sum_z [r^3E^3 + 3rEz^2 - \{(3n^2-7)/20\}\,rE] \\
&= r^4E^3 + \{3r^2(r^2-1)/12\}\,E - \{(3n^2-7)/20\}\,r^2E \\
&= r^4T_{3N}(E) - \{r^2(r^2-1)/10\}\,T_{1N}(E).
\end{aligned}$$

By similar calculations the following expressions are obtained for polynomials up to the fifth degree:

$$\sum_z T_{0n}(rE+z) = rT_{0N}(E);$$

$$\sum_z T_{1n}(rE+z) = r^2T_{1N}(E);$$

$$\sum_z T_{2n}(rE+z) = r^3T_{2N}(E);$$

$$\sum_z T_{3n}(rE+z) = r^4T_{3N}(E) - \{r^2(r^2-1)/10\}\,T_{1N}(E);$$

$$\sum_z T_{4n}(rE+z) = r^5T_{4N}(E) - \{3r^2(r^2-1)/7\}\,T_{2N}(E);$$

$$\sum_z T_{5n}(rE+z) = r^6T_{5N}(E) - \{10r^4(r^2-1)/9\}\,T_{3N}(E) - \{r^2(r^2-1)\,(n^2-6r^2+8)/126\}\,T_{1N}(E).$$

On substituting these expressions in equation (6), and using equation (3), the following expressions are obtained for the polynomial coefficients:

$$a_5 = r^{-6}A_5;$$

$$a_4 = r^{-5}A_4;$$

$$a_3 = r^{-4}[A_3 + \tfrac{10}{9}(1-r^{-2})\,A_5];$$

$$a_2 = r^{-3}[A_2 + \tfrac{3}{7}(1-r^{-2})\,A_4];$$

$$a_1 = r^{-2}[A_1 + \tfrac{1}{10}(1-r^{-2})\,A_3 + \tfrac{1}{126}(1-r^{-2})\,(N^2+8-6r^{-2})\,A_5];$$

$$a_0 = r^{-1}A_0.$$

In general
$$a_j = r^{-(j+1)} \sum_{k=j}^{p} \gamma_{jk} A_k, \tag{7}$$

where the values γ_{jj} are unity and γ_{jk} is zero for odd values of $j+k$. For polynomials of degrees up to the fourth the only non-zero coefficients other than γ_{jj} are γ_{13} and γ_{24}. These are listed in Table 1 for values of r from 2 to 15.

Table 1. *Values of* γ_{13} *and* γ_{24}

r	γ_{13}	γ_{24}	r	γ_{13}	γ_{24}
2	3/40	9/28	9	8/81	80/189
3	4/45	8/21	10	99/1000	297/700
4	3/32	45/112	11	12/121	360/847
5	12/125	72/175	12	143/1440	143/336
6	7/72	5/12	13	84/845	72/169
7	24/245	144/343	14	39/392	585/1372
8	63/640	27/64	15	112/1125	32/75

In tables of orthogonal polynomials the polynomial tabulated is
$$\xi'_j(\epsilon) = \lambda_j T_j(\epsilon).$$

Thus if
$$U_p(E) = \sum_j B_j \xi'_{jN}(E)$$

and
$$u_p(\epsilon) = \sum_j b_j \xi'_{jn}(\epsilon),$$

then
$$b_j = r^{-(j+1)} \sum_{k=j}^{p} \gamma_{jk} (\lambda_{kN}/\lambda_{jn}) B_k. \tag{8}$$

If the fitting is done by power moments, the coefficients A_j and a_j are obtained directly.

(2) *Standard errors of the coefficients*

If $\sigma(y)$ denotes the standard error of y and $\sigma(Y)$ the standard error of Y, it is clear that
$$\sigma^2(Y) = r\sigma^2(y) \tag{9}$$

and
$$\sigma^2(A_j) = \sigma^2(Y)/\Sigma T^2_{jN}. \tag{10}$$

From equation (7) it is apparent that $\sigma^2(a_j)$ will depend on p as well as j. However, provided N is not less than 10 the variation with p turns out to be insignificant. Considering the coefficient a_1 with $p = 5$,
$$\sigma^2(a_1) = r^{-4}[\sigma^2(A_1) + \tfrac{1}{100}(1-r^{-2})^2 \sigma^2(A_3) + \tfrac{1}{15876}(1-r^{-2})^2 (N^2+8-6r^{-2})^2 \sigma^2(A_5)].$$

On inserting the known expressions for ΣT^2_{jN} (Allan, 1930) in equation (10), it is found that
$$\sigma^2(a_1) \doteqdot r^{-4}\sigma^2(A_1)\,[1 + 6(1-r^{-2})^2 N^{-4}].$$

Similarly it may be shown that
$$\sigma^2(a_2) \doteqdot r^{-6}\sigma^2(A_2)\,[1 + 45(1-r^{-2})^2 N^{-4}],$$
$$\sigma^2(a_3) \doteqdot r^{-8}\sigma^2(A_3)\,[1 + 308(1-r^{-2})^2 N^{-4}].$$

Thus provided N is not less than 10 or so
$$\sigma^2(a_j) = r^{-2j+1}\sigma^2(A_j). \tag{11}$$

Let the values obtained for the coefficients by fitting a least-squares polynomial to the n observations be denoted by a'_j. Then

$$\sigma^2(a'_j) = \sigma^2(y)/\Sigma T^2_{jn},$$

and so the efficiency of the value a_j obtained from the grouped observations is

$$\eta_j = r^{2j+2}\frac{N(N^2-1)\dots(N^2-j^2)}{n(n^2-1)\dots(n^2-j^2)}\frac{\sigma^2(y)}{\sigma^2(Y)}$$

$$= \frac{(n^2-r^2)\dots(n^2-j^2r^2)}{(n^2-1)\dots(n^2-j^2)},$$

or $$\eta_j \doteqdot 1-\frac{j(j+1)(2j+1)}{6N^2}\{1-r^{-2}\}$$

$$+\frac{j(j^2-1)(2j-1)(10j^2+17j+6)}{360N^4}\{1-r^{-4}\}-\frac{\{j(j+1)(2j+1)\}^2}{36N^4}\{r^{-2}-r^{-4}\}. \quad (12)$$

From this expression the minimum number N of groups which must be retained for the efficiency to exceed a specified value can be determined. Table 2 gives the minimum number of groups required to ensure that the efficiencies do not fall below 0·98, 0·95, 0·90, 0·80 and 0·70. These values are calculated from (12) by neglecting terms r^{-k}. If r is small a slightly smaller number of groups would be permissible; as an example, for a_3 with $r = 2$, the value of N for an efficiency of 0·95 drops from 17 to 15.

Table 2. *Minimum values of N, the number of groups, to obtain stated efficiencies*

Degree j	1	2	3	4	5
Efficiency η_j					
0·98	8	16	27	39	53
0·95	5	10	17	25	33
0·90	4	8	12	18	24
0·80	3	5	9	12	16
0·70	—	4	7	10	13

(3) *Standard errors of the fitted values*

For the fitted values $$u_p(\epsilon) = \sum_j a_j T_j(\epsilon),$$

the standard errors may be found by expanding the coefficients a_j using equation (7), collecting terms in A_j, and applying equation (10). Thus, neglecting terms in r^{-2} in the expressions for γ_{jk},

$$\sigma^2[u_p(\epsilon)] = r^{-2}\sigma^2(A_0) + T^2_1(\epsilon)\,r^{-4}\sigma^2(A_1) + T^2_2(\epsilon)\,r^{-6}\sigma^2(A_2)$$
$$+[T_3(\epsilon)+\tfrac{1}{10}r^2T_1(\epsilon)]^2\,r^{-8}\sigma^2(A_3)$$
$$+[T_4(\epsilon)+\tfrac{3}{7}r^2T_2(\epsilon)]^2\,r^{-10}\sigma^2(A_4)$$
$$+[T_5(\epsilon)+\tfrac{10}{9}r^2T_3(\epsilon)+\tfrac{1}{126}r^2n^2T_1(\epsilon)]^2\,r^{-12}\sigma^2(A_5).$$

This is clearly a complicated function of ϵ. However, the leading terms $T_j(\epsilon)$ inside the square brackets will be much greater than the other terms except when $T_j(\epsilon)$ is small, and then the

contribution of the expression to the standard error is itself small. Hence to a reasonable approximation

$$\sigma^2[u_p(\epsilon)] \doteq \sum_j T_j^2(\epsilon)\,\sigma^2(a_j). \tag{13}$$

In the region of extrapolation the error $\sigma^2(a_p)$ becomes dominant, and the efficiency of the fitted value approaches the efficiency of the coefficient of highest degree. In the region of interpolation the efficiency will be somewhat greater than this value.

(4) *Estimation of the standard error of an observation*

The standard error of an observation may be estimated in several ways. Considering first the residuals $V_p(E)$ defined as

$$V_p(E) = Y(E) - \sum_j A_j T_{jN}(E),$$

it follows from least-squares theory that

$$\sum_E V_p^2(E) = \sum_E Y^2(E) - \sum_j A_j^2 \sum_E T_{jN}^2(E). \tag{14}$$

Hence

$$\begin{aligned}\exp\left[\Sigma\, V_p^2(E)\right] &= N\sigma^2(Y) - \sum_j \left\{\sum_E T_{jN}^2(E)\right\}\sigma^2(A_j)\\ &= (N-p-1)\,\sigma^2(Y)\\ &= r(N-p-1)\,\sigma^2(y),\end{aligned}$$

and the expression

$$S_p = \left\{\frac{\Sigma V_p^2(E)}{r(N-p-1)}\right\}^{\frac{1}{2}} \tag{15}$$

will provide an estimate of the standard error of an observation.

Another estimate may be obtained from the residuals $v_p(\epsilon)$ defined as

$$v_p(\epsilon) = y(\epsilon) - \sum_j a_j T_{jn}(\epsilon).$$

It can be shown (Guest, 1952, p. 255) that

$$\exp\left[\Sigma\, v_p^2(\epsilon)\right] = [n-(p+1)+\sum_j \{(1/\eta_j)-1\}]\,\sigma^2(y).$$

From equation (12), the last term in the square brackets is found to be approximately $p^4/12N^2$, and so it can be neglected. Hence

$$s_p = \left\{\frac{\Sigma v_p^2(\epsilon)}{n-p-1}\right\}^{\frac{1}{2}} \tag{16}$$

will provide an estimate of the standard error of an observation.

An indication of the relative efficiencies of the estimates s_p, S_p may be obtained by assuming that the residuals v_p are uncorrelated and of standard error $\sigma(y)$, and that the quantities $V_p/r^{\frac{1}{2}}$ are also uncorrelated and of standard error $\sigma(y)$. Both these assumptions are only very approximately true. Then, comparing (15) and (16), it will be seen that

$$\frac{(\text{S.E.}\, s_p)^2}{(\text{S.E.}\, S_p)^2} \doteq \frac{N}{n}.$$

That is, the relative efficiency of the estimate S_p is $1/r$. This estimate is very easily calculated, using equation (14), but is rather inefficient if r is large. On the other hand, there is no formula comparable with (14) by means of which s_p can be calculated, although the value

$$\Sigma y^2 - \Sigma a_j^2 \Sigma T_j^2(\epsilon)$$

could be used to give a rough approximation to Σv_p^2 if the efficiencies of the coefficients a_j are high.

(5) *The dropping of observations before grouping*

If the number of observations n is prime, it will be necessary to drop one or more observations at the ends of the range before grouping. The standard error of the least-squares coefficient a_j is proportional to $n^{-(j+\frac{1}{2})}$. Hence if the number of observations is reduced to n' by dropping ν observations before grouping, the efficiency of the estimate a_j will be reduced by the factor $(n'/n)^{2j+1}$. The efficiency of the grouping estimate is then the expression in equation (12) multiplied by the above factor. This factor may cause a considerable drop in efficiency. For example, if one observation is omitted when n is 50, the efficiency of a_3 is reduced by the factor 0·868. Table 3 shows the relative efficiencies for various values of n of the coefficients a_j when 1, 2 and 3 observations are omitted.

Table 3. *Relative efficiencies of a_j when ν observations are dropped from the original set of n observations*

	a_0			a_1			a_2		
ν	1	2	3	1	2	3	1	2	3
n									
20	0·950	0·900	0·850	0·857	0·729	0·614	0·774	0·590	0·444
30	0·967	0·933	0·900	0·903	0·813	0·729	0·844	0·708	0·590
40	0·975	0·950	0·925	0·927	0·857	0·791	0·881	0·774	0·677
50	0·980	0·960	0·940	0·941	0·885	0·831	0·904	0·815	0·734
75	0·987	0·973	0·960	0·961	0·922	0·885	0·935	0·874	0·815
100	0·990	0·980	0·970	0·970	0·941	0·913	0·951	0·904	0·859
150	0·993	0·987	0·980	0·980	0·961	0·941	0·967	0·935	0·904
250	0·996	0·992	0·988	0·988	0·976	0·964	0·980	0·961	0·941
500	0·998	0·996	0·994	0·994	0·988	0·982	0·990	0·980	0·970
	a_3			a_4			a_5		
ν	1	2	3	1	2	3	1	2	3
n									
20	0·698	0·478	0·321	0·630	0·387	0·232	0·569	0·314	0·167
30	0·789	0·617	0·478	0·737	0·537	0·387	0·689	0·468	0·314
40	0·838	0·698	0·579	0·796	0·630	0·496	0·757	0·569	0·424
50	0·868	0·751	0·648	0·834	0·693	0·573	0·801	0·638	0·506
75	0·910	0·828	0·751	0·886	0·784	0·693	0·863	0·743	0·638
100	0·932	0·868	0·808	0·914	0·834	0·760	0·895	0·801	0·715
150	0·954	0·910	0·868	0·942	0·886	0·834	0·929	0·863	0·801
250	0·972	0·945	0·919	0·965	0·930	0·897	0·957	0·915	0·876
500	0·986	0·972	0·959	0·982	0·965	0·947	0·978	0·957	0·936

The effect on the fitted values of dropping observations can be found by means of tables (Guest, 1950) of the functions $\rho_{p0}(k)$, where

$$k = 2\epsilon/n$$

and

$$\sigma^2[u_p(k)] = n^{-1}\rho_{p0}^2(k)\,\sigma^2(y).$$

The efficiency of the fitted value at the point ϵ will be given by

$$\eta[u_p(\epsilon)] = \sigma^2[u_p(k)]/\sigma^2[u_p(k')]$$
$$= \frac{n'}{n}\left(\frac{\rho_{p0}(k)}{\rho_{p0}(k')}\right)^2,$$

where (assuming the observations are omitted symmetrically from the two ends of the range)

$$k' = 2\epsilon/n', \qquad \nu \text{ even};$$
$$k' = (2\epsilon \pm 1)/n', \quad \nu \text{ odd}.$$

Tables and graphs of ρ_{p0} are given in the reference quoted. It is found that $\rho_{p0}(k)$ varies only slowly with k for a considerable part of the region of interpolation, and so in this part the efficiency of the fitted value is close to n'/n, the efficiency of the coefficient a_0. Beyond a value of $|k|$ given in Table 4, the variation of $\rho_{p0}(k)$ with k becomes very rapid and the efficiency drops sharply. For large values of $|k|$ the efficiency approaches that for the coefficient a_p.

Table 4. *Range of $|k|$ within which the efficiency of the fitted value when observations are omitted approximates to n'/n*

Degree p of polynomial	Range of $\lvert k \rvert$
2	0 to 0·50
3	0 to 0·70
4	0 to 0·80
5	0 to 0·85

Part II. Step-function methods

(1) *The straight line*

If $w_1(\epsilon)$ is any function for which $\quad \Sigma w_1(\epsilon) = 0, \qquad (1)$

then the equation $\quad \Sigma w_1(\epsilon)\,[y(\epsilon) - b_{11}\epsilon] = 0$

will provide an estimate $\quad b_{11} = \Sigma w_1(\epsilon)\, y(\epsilon)/\Sigma w_1(\epsilon)\,\epsilon \qquad (2)$

of the slope of the straight line which fits the observations. The standard error of this estimate is found from

$$\frac{\sigma^2(y)}{\sigma^2(b_{11})} = \frac{\{\Sigma w_1(\epsilon)\,\epsilon\}^2}{\Sigma w_1^2(\epsilon)}. \qquad (3)$$

The functions $w_1(\epsilon)$ which will be discussed in this part are the step functions. Step functions are functions which are constant in magnitude over specified ranges of the variable ϵ, the magnitude usually being different in different ranges. From equation (1) $w_1(\epsilon)$ will be an odd function of ϵ, and so, for any arbitrary function $f(\epsilon)$, $\Sigma w_1(\epsilon)\, f(\epsilon)$ will be of the form

$$\left[q_1\left(\sum_{0,\,\frac{1}{2}}^{\frac{1}{2}(n-1)} - \sum_{0,\,\frac{1}{2}}^{\frac{1}{2}(a_1-1)}\right) + q_2\left(\sum_{0,\,\frac{1}{2}}^{\frac{1}{2}(a_1-1)} - \sum_{0,\,\frac{1}{2}}^{\frac{1}{2}(a_2-1)}\right) + \ldots + q_m\left(\sum_{0,\,\frac{1}{2}}^{\frac{1}{2}(a_{m-1}-1)} - \sum_{0,\,\frac{1}{2}}^{\frac{1}{2}(a_m-1)}\right)\right][f(\epsilon) - f(-\epsilon)],$$

the numbers $\frac{1}{2}(a_j - 1)$ being the values of ϵ at the ends of the steps.

The number of steps, counting those corresponding to both positive and negative values of ϵ and the central step of zero weight, is $2m+1$. It is required to find, for any given m, the best values of the parameters a_j locating the steps and of the corresponding weights q_j.

The expression $\Sigma w_1(\epsilon)\,\epsilon$ will consist of terms of the form

$$2\sum_{0,\frac{1}{2}}^{\frac{1}{2}(a_j-1)}\epsilon = (a_j^2-1)/4.$$

If
$$\alpha_j = a_j/n, \tag{4}$$

then
$$\Sigma w_1(\epsilon)\,\epsilon = \frac{n^2}{4}[q_1(1-\alpha_1^2)+q_2(\alpha_1^2-\alpha_2^2)+\ldots+q_m(\alpha_{m-1}^2-\alpha_m^2)]. \tag{5}$$

Similarly
$$\Sigma w_1^2(\epsilon) = n[q_1^2(1-\alpha_1)+q_2^2(\alpha_1-\alpha_2)+\ldots+q_m^2(\alpha_{m-1}-\alpha_m)]. \tag{6}$$

Thus substituting these expressions in equation (3) and using the corresponding expression

$$n(n^2-1)/12 \doteqdot n^3/12$$

for the least-squares estimate, the efficiency of the estimate b_{11} is found to be

$$\eta(b_{11}) = \frac{3}{4}\frac{[q_1(1-\alpha_1^2)+q_2(\alpha_1^2-\alpha_2^2)+\ldots+q_m(\alpha_{m-1}^2-\alpha_m^2)]^2}{q_1^2(1-\alpha_1)+q_2^2(\alpha_1-\alpha_2)+\ldots+q_m^2(\alpha_{m-1}-\alpha_m)}. \tag{7}$$

Let the expression on the right be written as $\frac{3}{4}(N^2/D)$. To maximize $\eta(b_{11})$, the fraction N^2/D is differentiated with respect to the q_j, α_j, and the resultant expressions equated to zero. On differentiating with respect to q_j,

$$\alpha_{j-1}+\alpha_j = (N/D)\,q_j.$$

It is clear that the weights q_j may be multiplied by any arbitrary factor without altering $\eta(b_{11})$. It is convenient to choose the q_j so that (N/D) is unity. Then

$$\left.\begin{aligned}\alpha_{j-1}+\alpha_j &= q_j \quad (j=2 \text{ to } m);\\ 1+\alpha_1 &= q_1.\end{aligned}\right\} \tag{8}$$

On differentiating with respect to α_j,

$$\left.\begin{aligned}q_j+q_{j+1} &= 4\alpha_j \quad (j=1 \text{ to } m-1);\\ q_m &= 4\alpha_m.\end{aligned}\right\} \tag{9}$$

Combining equations (8) and (9), $\alpha_{j-1}+\alpha_{j+1} = 2\alpha_j$,

or
$$\alpha_{j-1}-\alpha_j = \Delta\alpha, \quad \text{a constant},$$
$$= 1-\alpha_1.$$

But
$$\alpha_{m-1}+\alpha_m = q_m = 4\alpha_m,$$

and so
$$\Delta\alpha = 2\alpha_m.$$

Thus
$$\alpha_m = \alpha_1-(m-1)\,\Delta\alpha = 1-m\Delta\alpha$$

and
$$\alpha_m = 1/(2m+1).$$

Hence
$$\alpha_j = \frac{2(m-j)+1}{2m+1} \tag{10}$$

and
$$q_j = \alpha_{j-1}+\alpha_j$$
$$= \frac{4(m-j+1)}{2m+1}. \tag{11}$$

The numerator N can be written as

$$N = \sum_{k=1}^{m} q_{m-k+1}(\alpha_{m-k}^2 - \alpha_{m-k+1}^2),$$

and so, using (10) and (11),

$$\begin{aligned} N &= \sum_k 4k\{(2k+1)^2-(2k-1)^2\}/(2m+1)^3 \\ &= 32\Sigma k^2/(2m+1)^3 \\ &= \frac{32m(m+1)}{6(2m+1)^2}. \end{aligned}$$

Thus, since $N = D$, the efficiency of the estimate b_{11} is

$$\eta(b_{11}) = \tfrac{3}{4}N = 1 - \frac{1}{(2m+1)^2}. \tag{12}$$

Hence for three steps the efficiency is 0·89, for five steps 0·96, and for seven steps 0·98.

Finally, it is convenient to remove the factor $4/(2m+1)$ from the weights, leaving

$$q_j = (m-j+1). \tag{13}$$

Thus for maximum efficiency the weights are $m, m-1, \ldots, 1$, and the values of the parameters locating the steps are

$$\frac{2m-1}{2m+1}, \quad \frac{2m-3}{2m+1}, \quad \ldots, \quad \frac{1}{2m+1}.$$

The number of observations in each step is

$$\tfrac{1}{2}(a_{j-1}-1) - \tfrac{1}{2}(a_j - 1) = \tfrac{1}{2}n(\alpha_{j-1} - \alpha_j) = n/(2m+1),$$

and the observations are divided uniformly among the $2m+1$ steps.

If the number of observations n is not an integral multiple of the number of steps, but

$$n = (2m+1)R + \rho \quad (|\rho| \leqslant m), \tag{14}$$

then the number of observations in some of the steps will be $R \pm 1$ instead of R. The methods of choosing the number n_j of observations in the step of weight j to give the maximum efficiency are listed in Table 5, together with the corresponding formulae for $\Sigma w_1(\epsilon)\epsilon$. For example, if $n = 66$ with 9 steps, the number of observations in the steps of largest $|\epsilon|$ is $n_4 = 7$, while $n_3 = 8, n_2 = 7, n_1 = 7, n_0 = 8$. In forming $\Sigma w_1(\epsilon) y(\epsilon)$, the sum of the n_j observations in the step of weight j corresponding to negative values of ϵ is subtracted from the sum of the n_j observations in the step of weight j corresponding to positive values of ϵ, and the difference is multiplied by the weight j. The value of b_{11} is then given by equation (2).

(2) *Polynomials of higher degree*

(a) *Second degree polynomials*

To obtain an estimate b_{22} of the second degree coefficient a function $w_2(\epsilon)$ is required such that

$$\Sigma w_2(\epsilon) = 0, \tag{15}$$

$$\Sigma w_2(\epsilon)\epsilon = 0. \tag{16}$$

The estimate is then

$$b_{22} = \Sigma w_2(\epsilon) y(\epsilon)/\Sigma w_2(\epsilon)\epsilon^2, \tag{17}$$

and its standard error is found from

$$\frac{\sigma^2(y)}{\sigma^2(b_{22})} = \frac{\{\Sigma w_2(\epsilon)\,\epsilon^2\}^2}{\Sigma w_2^2(\epsilon)}. \tag{18}$$

Condition (16) is satisfied if $w_2(\epsilon)$ is an even function of ϵ. Condition (15) requires that the values $w_2(\epsilon)$ should fall into two groups, one group in which $w_2(\epsilon)$ is positive and the other group in which $w_2(\epsilon)$ is negative. Accordingly, the appropriate step functions will be such that, for any arbitrary function $f(\epsilon)$, $\Sigma w_2(\epsilon)\,f(\epsilon)$ is of the form

$$\left[\left\{q_1\left(\sum_{-\frac{1}{2}(n-1)}^{\frac{1}{2}(n-1)} - \sum_{-\frac{1}{2}(a_1-1)}^{\frac{1}{2}(a_1-1)}\right) + q_2\left(\sum_{-\frac{1}{2}(a_1-1)}^{\frac{1}{2}(a_1-1)} - \sum_{-\frac{1}{2}(a_2-1)}^{\frac{1}{2}(a_2-1)}\right) + \ldots\right\}\right.$$
$$\left. - \left\{r_1\left(\sum_{-\frac{1}{2}(b_1-1)}^{\frac{1}{2}(b_1-1)} - \sum_{-\frac{1}{2}(b_2-1)}^{\frac{1}{2}(b_2-1)}\right) + r_2\left(\sum_{-\frac{1}{2}(b_2-1)}^{\frac{1}{2}(b_2-1)} - \sum_{-\frac{1}{2}(b_3-1)}^{\frac{1}{2}(b_3-1)}\right) + \ldots\right\}\right][f(\epsilon)].$$

The functions will be called single-step functions if there is one step in each group, double-step functions if there are two steps.

Table 5. *Step functions for the fitting of straight lines*

(n_j: number of observations in step of weight j. $n = (2m+1)R+\rho$: total number of observations.)

Three steps: $n = 3R+\rho$

ρ	n_1	n_0	$\Sigma w_1(\epsilon)\,\epsilon$
0	R	R	$2R^2$
± 1	R	$R \pm 1$	$2R^2 \pm R$

$\eta(b_{11}) = 0{\cdot}889.$

Five steps: $n = 5R+\rho$

ρ	n_2	n_1	n_0	$\Sigma w_1(\epsilon)\,\epsilon$
0	R	R	R	$10R^2$
± 1	R	R	$R \pm 1$	$10R^2 \pm 3R$
± 2	R	$R \pm 1$	R	$10R^2 \pm 7R+1$

$\eta(b_{11}) = 0{\cdot}960.$

Seven steps: $n = 7R+\rho$

ρ	n_3	n_2	n_1	n_0	$\Sigma w_1(\epsilon)\,\epsilon$
0	R	R	R	R	$28R^2$
± 1	R	R	R	$R \pm 1$	$28R^2 \pm 6R$
± 2	R	R	$R \pm 1$	R	$28R^2 \pm 13R+1$
± 3	R	$R \pm 1$	R	$R \pm 1$	$28R^2 \pm 22R+4$

$\eta(b_{11}) = 0{\cdot}980.$

Nine steps: $n = 9R+\rho$

ρ	n_4	n_3	n_2	n_1	n_0	$\Sigma w_1(\epsilon)\,\epsilon$
0	R	R	R	R	R	$60R^2$
± 1	R	R	R	R	$R \pm 1$	$60R^2 \pm 10R$
± 2	R	R	R	$R \pm 1$	R	$60R^2 \pm 21R+1$
± 3	R	$R \pm 1$	R	R	$R \pm 1$	$60R^2 \pm 39R+6$
± 4	R	$R \pm 1$	R	$R \pm 1$	R	$60R^2 \pm 50R+10$

$\eta(b_{11}) = 0{\cdot}988.$

The expression $\Sigma w_2(\epsilon)\,\epsilon^2$ will contain terms

$$\sum_{-\frac{1}{2}(a_j-1)}^{\frac{1}{2}(a_j-1)} \epsilon^2 = a_j(a_j^2-1)/12,$$

for which the approximation $n^3\alpha_j^3/12$ will be used, where α_j is a_j/n. Hence

$$\Sigma w_2(\epsilon)\,\epsilon^2 = \frac{n^3}{12}[\{q_1(1-\alpha_1^3)+q_2(\alpha_1^3-\alpha_2^3)+\ldots\}-\{r_1(\beta_1^3-\beta_2^3)+\ldots+r_m\beta_m^3\}],$$

$$\Sigma w_2^2(\epsilon) = n\,[\{q_1^2(1-\alpha_1)+q_2^2(\alpha_1-\alpha_2)+\ldots\}+\{r_1^2(\beta_1-\beta_2)+\ldots+r_m^2\beta_m\}].$$

Substituting these expressions in equation (18) and using the corresponding expression

$$n(n^2-1)\,(n^2-4)/180 \doteqdot n^5/180$$

for the least-squares estimate, the efficiency of the estimate b_{22} is found to be

$$\eta(b_{22}) = \frac{5}{4}\frac{[\{q_1(1-\alpha_1^3)+\ldots\}-\{r_1(\beta_1^3-\beta_2^3)+\ldots\}]^2}{\{q_1^2(1-\alpha_1)+\ldots\}+\{r_1^2(\beta_1-\beta_2)+\ldots\}}, \tag{19}$$

with the conditional equation

$$\Sigma w_2(\epsilon) = \{q_1(1-\alpha_1)+\ldots\}-\{r_1(\beta_1-\beta_2)+\ldots\} = 0. \tag{20}$$

The values q, α, r, β, for maximum efficiency, were found by the method of undetermined multipliers. The resulting equations are simultaneous quadratics, and the values of the parameters were obtained by successive approximations. The solutions for single-step and double-step functions are listed in Table 6.

The maximum efficiency of the estimate b_{22} is found to be 0·8958 for single-step functions and 0·9630 for double-step functions.

Table 6. *Values of parameters for maximum efficiencies*

Second degree coefficient			Third degree coefficient		
Single-step functions		Double-step functions	Single-step functions		Double-step functions
α	0·7363	α_1 0·8482 α_2 0·6757	α	0·8621	α_1 0·9222 α_2 0·8309
β	0·4414	β_1 0·5007 β_2 0·3207	β_1 β_2	0·7024 0·1205	β_1 0·7399 β_2 0·6536 β_3 0·1265
q	1·3042	q_1 1·5774 q_2 0·7590	q	1·0393	q_1 1·2465 q_2 0·5940
r	0·7793	r_1 0·4762 r_2 0·8875	r	0·5573	r_1 0·3167 r_2 0·5919
η_2	0·8958	η_2 0·9630	η_3	0·9014	η_3 0·9473

(*b*) *Third degree polynomials*

To obtain an estimate b_{33} of the third degree coefficient a function $w_3(\epsilon)$ is required such that

$$\Sigma w_3(\epsilon) = \Sigma w_3(\epsilon)\,\epsilon^2 = 0, \tag{21}$$

$$\Sigma w_3(\epsilon)\,\epsilon = 0. \tag{22}$$

The estimate is then

$$b_{33} = \Sigma w_3(\epsilon)\,y(\epsilon)/\Sigma w_3(\epsilon)\,\epsilon^3, \tag{23}$$

and its standard error is found from

$$\frac{\sigma^2(y)}{\sigma^2(b_{33})} = \frac{\{\Sigma w_3(\epsilon)\,\epsilon^3\}^2}{\Sigma w_3^2(\epsilon)}. \tag{24}$$

Condition (21) requires that $w_3(\epsilon)$ be an odd function of ϵ. Proceeding as in § (*a*), the efficiency of the estimate is found to be given by the expression

$$\eta(b_{33}) = \frac{175}{64}\frac{[\{q_1(1-\alpha_1^4)+\ldots\}-\{r_1(\beta_1^4-\beta_2^4)+\ldots+r_m(\beta_m^4-\beta_{m+1}^4)\}]^2}{\{q_1^2(1-\alpha_1)+\ldots\}+\{r_1^2(\beta_1-\beta_2)+\ldots+r_m^2(\beta_m-\beta_{m+1})\}}, \tag{25}$$

with the conditional equation

$$\Sigma w_3(\epsilon)\,\epsilon = \{q_1(1-\alpha_1^2)+\ldots\}-\{r_1(\beta_1^2-\beta_2^2)+\ldots\} = 0. \tag{26}$$

The maximum efficiencies for single-step functions and for double-step functions are found to occur at the values of the parameters listed in Table 6. The maximum efficiency for the estimate b_{33} is 0·9014 for single-step functions and 0·9473 for double-step functions.

(3) *Polynomial coefficients*

The estimate of the coefficient of ϵ^j in the power series expansion of the polynomial of degree p will be denoted by b_{pj}. b_{pj} can be expressed as a linear function of the quantities

$$b_{kk} = \Sigma w_k(\epsilon)\,y(\epsilon)/\Sigma w_k(\epsilon)\,\epsilon^k,$$

where

$$\Sigma w_k(\epsilon)\,\epsilon^m = 0 \quad (k>m).$$

For if the estimates are to be unbiased

$$\Sigma w_j(\epsilon)\,[y(\epsilon)-\Sigma b_{pk}\epsilon^k] = 0,$$

and hence

$$\Sigma w_j(\epsilon)\,y(\epsilon) = \Sigma w_j(\epsilon)\sum_{k=j}^{p} b_{pk}\epsilon^k.$$

Dividing by $\Sigma w_j(\epsilon)\,\epsilon^j$,

$$b_{jj} = b_{pj} - \sum_{j+1}^{p}\alpha_{kj}b_{pk},$$

where

$$\alpha_{kj} = -\Sigma w_j(\epsilon)\,\epsilon^k/\Sigma w_j(\epsilon)\,\epsilon^j. \tag{27}$$

Hence

$$b_{pj} = \sum_{k=j}^{p}\beta_{kj}b_{kk}, \tag{28}$$

where

$$\left.\begin{aligned}\beta_{kj} &= \alpha_{kj} + \sum_{j+1}^{k-1}\alpha_{km}\beta_{mj},\\ \beta_{kk} &= 1,\end{aligned}\right\} \tag{29}$$

and β_{kj} vanishes if $j+k$ is odd. For the second and third degree polynomials

$$b_{20},\quad b_{30} = b_{00}+\beta_{20}b_{22},\quad b_{32} = b_{22},$$
$$b_{31} = b_{11}+\beta_{31}b_{33},\quad b_{21} = b_{11},$$

where

$$\beta_{20} = -\Sigma\epsilon^2/n = -(n^2-1)/12,$$
$$\beta_{31} = -\Sigma w_1(\epsilon)\,\epsilon^3/\Sigma w_1(\epsilon)\,\epsilon.$$

(4) *The fitted values*

The fitted polynomial of the third degree is

$$u_3(\epsilon) = b_{30} + b_{31}\epsilon + b_{32}\epsilon^2 + b_{33}\epsilon^3$$
$$= b_{00} + b_{11}\epsilon + b_{22}(\epsilon^2 + \beta_{20}) + b_{33}(\epsilon^3 + \beta_{31}\epsilon).$$

The coefficients b_{jj} are independent, except for b_{11} and b_{33}, where

$$\operatorname{cov}(b_{11}b_{33})/\sigma^2(y) = \Sigma w_1(\epsilon)\, w_3(\epsilon)/\Sigma w_1(\epsilon)\, \epsilon \Sigma w_3(\epsilon)\, \epsilon^3.$$

Hence the standard error of the fitted value is obtained from

$$\sigma^2[u_3(\epsilon)] = \sigma^2(b_{00}) + \epsilon^2\sigma^2(b_{11}) + (\epsilon^2 + \beta_{20})^2\sigma^2(b_{22}) + (\epsilon^3 + \beta_{31}\epsilon)^2\sigma^2(b_{33})$$
$$+ 2(\epsilon^4 + \beta_{31}\epsilon^2)\operatorname{cov}(b_{11}b_{33}).$$

The standard error of the fitted value for polynomials of lower degree is obtained by merely dropping the higher degree terms.

It is convenient to introduce the variable $k = 2\epsilon/n$. Then, neglecting terms of order n^{-2} and writing η_j for $\eta(b_{jj})$,

$$\frac{\sigma^2[u_3(k)]}{\sigma^2(y)} = \frac{1}{n}\left[1 + \frac{3}{\eta_1}k^2 + \frac{5}{4\eta_2}(3k^2-1)^2 + \frac{175}{4\eta_3}(k^3 + 4\beta_{31}n^{-2}k)^2 + \tfrac{1}{8}(k^4 + 4\beta_{31}n^{-2}k^2)\frac{\operatorname{cov}(b_{11}b_{33})}{\sigma^2(y)}\right]. \tag{30}$$

By calculating the value of this expression for the least-squares and step-function polynomials, the efficiencies of the fitted values may be obtained for various k.

The efficiencies of the fitted values with single-step and double-step functions are shown as functions of k for polynomials of the first, second and third degrees in Table 7.

Table 7. *Efficiencies of the fitted values using step-functions* ($k = 2\epsilon/n$)

Degree	$p = 1$		$p = 2$		$p = 3$	
$\lvert k \rvert$	Single-step	Double-step	Single-step	Double-step	Single-step	Double-step
0	1·000	1·000	0·939	0·979	0·939	0·979
0·2	0·987	0·996	0·942	0·980	0·941	0·973
0·4	0·961	0·987	0·948	0·982	0·948	0·967
0·6	0·939	0·979	0·939	0·979	0·956	0·970
0·8	0·924	0·973	0·916	0·971	0·919	0·972
1·0	0·914	0·970	0·904	0·966	0·887	0·959
1·2	0·908	0·967	0·899	0·964	0·888	0·953
1·4	0·903	0·966	0·897	0·963	0·891	0·950
1·6	0·900	0·964	0·897	0·963	0·894	0·949

(5) *Tabulation of step functions*

In a previous paper (Guest, 1951) the optimum single-step functions have been listed for the range $p = 1\,(1)\,5$, $n = 7\,(1)\,55$. These tables were obtained by calculating the efficiencies for chosen arrangements of groups at each value of n and selecting the method of grouping which gave maximum efficiency. In point of fact such a procedure is unnecessary except when n is small. The efficiency decreases only slowly as the arrangement of groups departs

from the optimum arrangement, and it is quite satisfactory to use the general parameters α to determine the number of observations in each group.

In tabulating the double-step functions the location of the groups can be obtained from the parameters α, β. The corresponding weights q, r are not exact multiples of one another and so the optimum weights will not be integers. It is therefore desirable to take values slightly different from the optimum weights. The recommended weights are:

2nd degree: $q_1 = 10w, \quad q_2 = 5w, \quad r_1 = 3w,$
w, r_2 integers adjusted so that $\Sigma w_2(\epsilon) = 0;$

3rd degree: $q_1 = 4w, \quad q_2 = 2w, \quad r_1 = w,$
w, r_2 integers adjusted so that $\Sigma w_3(\epsilon)\,\epsilon = 0.$

Then $\eta_2 = 0{\cdot}9628$, $\eta_3 = 0{\cdot}9471$, and the drop in efficiency is negligible in each case. By this choice of weights the number of weights to be tabulated is reduced to two (w and r_2).

Single-step functions give efficiencies of about 0·90. With double-step functions, for a second degree polynomial the efficiencies of all quantities are greater than 0·96. For a third degree polynomial the efficiency of the fitted value is greater than 0·96 in the region of interpolation. The efficiency of the fitted value in the region of extrapolation drops to about 0·945, the efficiency of the coefficient b_{33}.

Tables of both single-step and double-step functions have been prepared for values of n up to 100, the single-step functions up to the fifth degree and the double-step functions up to the third degree. These tables are too long to be reproduced here, but copies may be obtained from the author.

CONCLUSION

An attempt has been made to assess the saving in time that would eventuate in practical examples from the use of grouping methods. The example chosen for the test was that dealing with coded sugar prices which was used by Anderson & Houseman (1942). Table 8 gives a summary of the solutions obtained by the various methods, and the time taken using each method (including the time required for the checking of all calculations). The calculations were all carried out by a computor familiar with the methods, using an electric machine equipped with semi-automatic multiplication.

Table 8. *Solutions of example on coded sugar prices obtained by different methods*

	Least-squares methods			Step-function methods	
	62 observations	31 groups	15 groups	Single-step	Double-step
$-b_{33} \times 10^3$	$2{\cdot}59 \pm 0{\cdot}32$	2·59	2·95	2·51	2·44
$b_{32} \times 10^2$	$2{\cdot}86 \pm 0{\cdot}50$	2·85	2·92	2·92	2·84
$b_{31} \times 10$	$9{\cdot}8 \pm 2{\cdot}0$	9·8	12·4	10·1	8·7
b_{30}	$12{\cdot}4 \pm 2{\cdot}2$	12·4	11·7	12·2	12·4
Minutes required:					
By machine	43	31	29	20	28
By logarithms	—	—	—	41	62

An examination of the times required for the calculation of the coefficients shows that the saving in time through the use of grouping methods is not as large as might be expected. The reduction in the strain on the computor is more significant than might be deduced from the times shown, because with grouping methods the multiplying factors are both smaller in number and smaller in size, and the chance of making an error is much reduced. In fact the step-function methods are almost foolproof, except for the possibility of copying errors. The saving in time becomes much more pronounced if computational facilities are limited—for instance, if only a hand-operated machine is available. Step-function methods can be used even when no machine is available, the quantities $\Sigma w_j(\epsilon)\, y(\epsilon)$ being obtained by simple addition and the coefficients b_{pj} by the use of logarithms.

It will be seen that the values obtained by using groups of two are the same as the least-squares values for the original sixty-two observations. However, if the observations are grouped in fours, the two end observations must be omitted and the values are noticeably different. The fact that the deviations for the coefficients b_{31}, b_{33} exceed the standard errors indicates that the variations from a smooth curve are not random, as is apparent from the nature of the data. The deviations for the step-function solutions are all less than the standard errors.

The weakness of the step-function methods lies in the difficulty of estimating the standard errors, and of deciding the degree of the polynomial to be used if this is not already known. This is because there is no formula comparable with the

$$\Sigma v_p^2 = \Sigma y^2 - \sum_j a_j^2 \Sigma T_j^2$$

of the least-squares method. To estimate the standard error of an observation it is necessary to calculate the residuals individually. If the fitted values at the points of observation are also required the calculation of the residuals will not take much time, but otherwise the time spent in calculating the residuals will largely offset the time saved in the calculation of the coefficients b_{pj}.

In practice by far the largest number of curves to be fitted are of the first degree. Step-function methods of fitting straight lines are especially useful in undergraduate instruction where calculating machines are not usually available, and they may also be used to give a rapid independent check on the least-squares solution. The loss in accuracy compared with the least-squares method is quite small, and the use of single-step and double-step functions in teaching and in routine work is consequently strongly recommended.

If the observations are to be grouped before fitting a least-squares curve it is important that the number of observations omitted at the ends of the range be very small. Tables 2 and 3 may be used as a guide in selecting the number of groups to give a satisfactory value for the efficiency.

REFERENCES

Allan, F. E. (1930). *Proc. Roy. Soc. Edinb.* **50**, 310.

Anderson, R. L. & Houseman, E. E. (1942). *Res. Bull. Ia Agric. Exp. Sta.* no. 292.

Guest, P. G. (1950). *Aust. J. Sci. Res.* A, **3**, 173.

Guest, P. G. (1951). *Ann. Math. Statist.* **22**, 537.

Guest, P. G. (1952). *Aust. J. Sci. Res.* A, **5**, 238.

Guest, P. G. (1953). *J. Math. Phys.* **32**, 68.

NEW TECHNIQUES FOR THE ANALYSIS OF ABSENTEEISM DATA

By A. G. ARBOUS and H. S. SICHEL

National Institute for Personnel Research, South African Council for Scientific and Industrial Research

1. Introduction

The shortage of competent man-power has in recent years accentuated the problem of absence in industry, and revived both managerial and scientific interest in the subject. To management these occurrences are a drain on the productivity of the concern which can no longer be dealt with by methods which, presumably, were effective when the supply of labour was abundant. They now turn for advice to the social scientist, who, looking deeper into the problem, is inclined to regard absenteeism as symtomatic of occupational casualty; undesirable for the well-being of the persons engaged in production.

In view of this state of affairs it is surprising that the statistical tools for analysing and interpreting absence data should have remained relatively undeveloped. By contrast one compares the more thorough, though not necessarily more successful, attack which has been made in the field of accidents. And yet, as will be seen later, the two phenomena have much in common. In the past the analysis of absences has confined itself largely to the monotonous and often meaningless calculation of a series of 'rates' (frequency, severity and disability) for innumerable classifications of the data. These 'rates' are not amenable to statistical analysis, and their interpretation is consequently difficult. More recently authors such as Fox & Scott (1943) have departed from this approach by giving frequency distributions of absences and time lost, without attempting to give these any mathematical description. Further advance was made by Russell, Whitwell & Ryle (1947) and Sutherland & Whitwell (1948), who applied to absence distributions the models developed by Greenwood & Yule (1920) in the study of accidents, and claim thereby to have established the phenomenon of absence-proneness. However, the inadequacy of inferences drawn from the fitting of theoretical univariate distributions to observed sets of data has been clearly demonstrated by Maritz (1950) and Arbous & Kerrich (1951), and the case would appear to be unsubstantiated.

After the present study was completed we find that another investigator (Norris, 1951) had adopted the approach suggested by Lundberg (1940) and Maritz (1950), viz. that the existence of proneness in absence behaviour can only be established by the method of correlation. Unfortunately, apart from showing the existence of correlation, its physical significance was not demonstrated—a defect which we hope to remedy in this study.

It is therefore the object of this paper to develop more precisely the mathematical model for describing the distribution of absences to a group of people in single- and double-exposure periods, thus allowing for both the univariate and bivariate case. We shall, moreover, consider the physical significance of the properties of such distributions (*a*) for the interpretation of research findings, with particular reference to the phenomenon of absence-proneness, and (*b*) for the development of practical techniques. For reasons, which cannot be elaborated here, we shall confine our observations to absences which occur (for various reasons) to

individuals, and not the duration of absence. One absence is thus defined as a period (of any length) of unbroken non-attendance at work. These absences can be grouped according to various causes, if it is desired.

2. The mathematical model

(a) *The univariate negative binomial*

Let the probability distribution of absences for an individual during an exposure period of length kt be represented by

$$\text{prob.}\,(x) = e^{-k\lambda}\frac{(k\lambda)^x}{x!}. \tag{1}$$

Following the original suggestion of Greenwood & Yule (1920) we shall assume that in a group of persons the individual liabilities, as measured by the λ's, are distributed according to

$$dF = \left(\frac{p}{m}\right)^p \frac{1}{\Gamma(p)} e^{-(p/m)\lambda}\lambda^{p-1}d\lambda \quad (0 \leqslant \lambda \leqslant \infty), \tag{2}$$

where m and p are the unknown parameters of the law. For (2) we have

$$E(\lambda) = m, \tag{3}$$

$$\sigma_\lambda = \frac{m}{\sqrt{p}}, \tag{4}$$

$$\beta_1(\lambda) = \frac{4}{p}. \tag{5}$$

The probability of finding a person in the group having x absences in one and the same exposure period of duration kt is given by

$$f(x) = \left(\frac{p}{m}\right)^p \frac{1}{\Gamma(p)}\frac{k^x}{\Gamma(x+1)}\int_0^\infty \lambda^{x+p-1}e^{-(k+p/m)\lambda}d\lambda. \tag{6}$$

On integration we find

$$f(x) = \left(\frac{p}{p+km}\right)^p \frac{1}{\Gamma(p)}\frac{\Gamma(x+p)}{\Gamma(x+1)}\left(\frac{km}{p+km}\right)^x, \tag{7}$$

which is the negative binomial distribution with exponent p and mean km. For an observational period of length t we have $k = 1$ and the above distribution takes on the standard form. From equation (7) we deduce:

(1) The mean number of absences incurred by a group of persons in an observational period is directly proportional to the length of that period, i.e.

$$\mu_1' = km, \tag{8}$$

where m is the mean for duration t and km the mean for duration kt.

(2) From the moments of the negative binomial we find for the variance of absences in period kt

$$\mu_2 = k^2\left(\frac{m}{k} + \frac{m^2}{p}\right). \tag{9}$$

(3) Parameter p is independent of the length of the exposure.

From the first four moments of the standard negative binomial distribution ($k = 1$) we find the Pearsonian coefficients of skewness and kurtosis

$$\beta_1 = \frac{(p+2m)^2}{mp(p+m)} \tag{10}$$

and

$$\beta_2 = 3 + \frac{6}{p} + \frac{p}{m(p+m)}. \tag{11}$$

For p large and m finite

$$\beta_1 \doteqdot \frac{1}{m}, \quad \beta_2 \doteqdot 3+\frac{1}{m},$$

which are the respective shape coefficients of the Poisson distribution. For both p and m large, $\beta_1 \doteqdot 0$, $\beta_2 \doteqdot 3$, and under these conditions the negative binomial tends towards the normal distribution.

(b) The symmetrical bivariate negative binomial

If the existence or non-existence of absence-proneness in a group of individuals is to be established by correlational methods (Maritz, 1950) we must construct a bivariate model. The bivariate negative binomial distribution was first given by Lundberg (1940) and has again been derived recently by Arbous & Kerrich (1951). In the field of absenteeism it is convenient to split the total exposure period into two equal parts. In what follows we shall deduce the symmetrical bivariate negative binomial which is a specialized but simpler case of Lundberg's distribution. On the other hand, we believe that some of the formulae pertaining to the arrays, the limiting conditions of correlation and the operating characteristic are new.

Suppose that we deal with two equilong non-overlapping periods of exposure. Each individual will have x absences in the first and y absences in the second period. Provided an individual's absences in the second period are independent of those incurred in the first, and provided the individual liabilities to absence follow equation (2), we find for the bivariate symmetrical binomial

$$\phi(y,x) = \left(\frac{p}{m}\right)^p \frac{1}{\Gamma(p)} \frac{1}{\Gamma(y+1)\,\Gamma(x+1)} \int_0^\infty \lambda^{y+x+p-1} e^{-(2+p/m)\lambda}\, d\lambda. \tag{12}$$

After integration

$$\phi(y,x) = \left(\frac{p}{p+2m}\right)^p \frac{1}{\Gamma(p)} \frac{\Gamma(y+x+p)}{\Gamma(y+1)\,\Gamma(x+1)} \left(\frac{m}{p+2m}\right)^{y+x}. \tag{13}$$

The marginal distributions of (13), their means and their variances are given by equations (7), (8) and (9) respectively with $k = 1$. The product-moment correlation of a symmetrical bivariate distribution is

$$\rho_{yx} = \frac{\mu'_{11} - (\mu'_1)^2}{\mu_2},$$

where μ'_{11} is the product moment about the origin. This may be found from

$$\mu'_{11} = \sum_{y=0}^{\infty}\sum_{x=0}^{\infty} yx\phi(y,x) = \left(\frac{p}{p+2m}\right)^p \frac{1}{\Gamma(p)} \sum_{y=1}^{\infty}\sum_{x=1}^{\infty} \frac{\Gamma(y+x+p)}{\Gamma(y)\,\Gamma(x)} \left(\frac{m}{p+2m}\right)^{y+x}. \tag{14}$$

Substitute

$$y = y'+1, \tag{15a}$$

$$x = x'+1, \tag{15b}$$

$$p = p'-2, \tag{15c}$$

$$m = \left(1-\frac{2}{p'}\right)m', \tag{15d}$$

into the double sum on the right-hand side of equation (14) and we find for it

$$\left(\frac{m}{p+2m}\right)^2 \sum_{y'=0}^{\infty}\sum_{x'=0}^{\infty} \frac{\Gamma(y'+x'+p')}{\Gamma(y'+1)\,\Gamma(x'+1)} \left(\frac{m'}{p'+2m'}\right)^{y'+x'} = \left(\frac{m}{p+2m}\right)^2 \Gamma(p') \left(\frac{p'+2m'}{p'}\right)^{p'}$$

$$= m^2\Gamma(p)\left(\frac{p+1}{p}\right)\left(\frac{p+2m}{p}\right)^p.$$

Finally, by substitution of the last expression into equation (14)

$$\mu'_{11} = \left(\frac{p+1}{p}\right) m^2 \tag{16}$$

and

$$\rho_{yx} = \frac{m}{p+m}. \tag{17}$$

From (17) we arrive at the important conclusion that, given a phenomenon which may be described by a symmetrical bivariate negative binomial, the product-moment correlation may be expressed in terms of the parameters of the univariate marginal distribution. This result was already known to Newbold (1927).

If we interpret ρ_{yx} as a reliability coefficient of observed absences in one period of duration t and we lengthen this period to kt we step up the reliability to

$$\rho_{yx,kt} = \frac{km}{p+km}, \tag{18}$$

in view of equation (7). We could have deduced this result from the general Spearman-Brown reliability formula

$$\rho_{yx,k} = \frac{k\rho_{yx}}{1+(k-1)\rho_{yx}},$$

because substitution of (17) into the above gives again (18).

We now derive the probability distributions of arrays for $x = 0, 1, 2, \ldots, \infty$

$$\gamma(y|x) = \frac{\phi(y|x)}{f(x)},$$

from which we find the array distribution

$$\gamma(y|x) = \left(1-\frac{m}{p+2m}\right)^{x+p} \frac{1}{\Gamma(x+p)} \frac{\Gamma(y+x+p)}{\Gamma(y+1)} \left(\frac{m}{p+2m}\right)^{y}. \tag{19}$$

Substitution of

$$x+p = p'', \tag{20a}$$

$$m = \frac{(p''-x)\,m''}{p''-m''}, \tag{20b}$$

into (19) gives

$$\gamma(y|x) = \left(\frac{p''}{p''+m''}\right)^{p''} \frac{1}{\Gamma(p'')} \frac{\Gamma(y+p'')}{\Gamma(y+1)} \left(\frac{m''}{p''+m''}\right)^{y}, \tag{21}$$

which is a standard negative binomial with mean and variance (in the original parameters)

$$\mu'_{1.\,\text{Arr.}} = \frac{m}{p+m}(x+p) = \rho x + \rho p,$$

$$\mu_{2.\,\text{Arr.}} = \frac{(p+2m)\,m}{(p+m)^2}(x+p) = (1+\rho)(\rho x+\rho p).$$

Hence for the regression y on x

$$\tilde{y} = \rho x + \rho p, \tag{22}$$

and the variance within an array

$$\mu_{2.\,\text{Arr.}} = (1+\rho)\,\mu'_{1.\,\text{Arr.}}, \tag{23}$$

From the last two equations, it follows that the regressions y on x and for reason of symmetry also x on y, are linear but heteroscedastic. The variances within arrays are increasing linearly with the array means. The slope constant of the regression lines is equal to the product-moment correlation ρ.

It is interesting to note that for the limiting correlation $\rho \to 1$ the bivariate negative binomial distribution does not tend towards a concentration along the regression line as is the case with the bivariate normal distribution. In a strict mathematical sense the limiting process may be obtained in two ways:

$$(1)\ \lim_{p\to 0}\rho = \lim_{p\to 0}\frac{m}{p+m} = 1, \qquad \text{for finite } m.$$

$$(2)\ \lim_{m\to\infty}\rho = \lim_{m\to\infty}\frac{m}{p+m} = 1, \quad \text{for finite } p.$$

In the first case the regression line will tend towards

$$\tilde{y} = x, \tag{24}$$

but the variance within arrays towards

$$\mu_{2.\,\text{Arr.}} = 2x, \tag{25}$$

which is far from zero spread. The vast majority of observations will be concentrated in the $(0, 0)$ cell with the remainder thinly spread over a wide range of absences.

In the second case the regression line will tend towards

$$\tilde{y} = x + p. \tag{26}$$

The slope of this line (45°) is again the same as in the first case, but a constant has to be added. The array variance becomes

$$\mu_{2.\,\text{Arr.}} = 2x + 2p, \tag{27}$$

which is in fact maximal for $\rho = 1$.

From a practical point of view the distinction between the two limiting processes is of little consequence with respect to the regression lines. As $m \to \infty$ the whole bivariate distribution will shift away from the $(0, 0)$ corner, and thus to get it on to a frequency table for inspection, we must reduce the length on the paper represented by a unit of x or y; e.g. by dividing all observations by m. In this case the difference between equations (24) and (26) would be negligible as p/m would be exceedingly small.

On the other hand, in practice the difference between the two limiting processes may be significant, depending upon whether one looks at array variability from the absolute or relative point of view. In the former case, which will be needed when predicting future absences for a given person, as $p \to 0$ and consequently $\rho \to 1$, the standard error of prediction for a fixed x will first increase, pass through a maximum and then decrease. For the second limiting process, where $m \to \infty$ and $\rho \to 1$, the standard error of prediction for fixed x will monotonically increase reaching a maximum at $\rho = 1$. This result is unusual in the sense that when predicting future absences for two different persons, with the same x and exposure periods, from two different groups, the standard error of prediction is smaller in respect of the one who comes from a lower correlation surface, and vice versa.

In the latter case, when considering relative variability, the behaviour of the bivariate negative binomial surface does not seem to be anomalous. In this case the value of the correlation from the point of view of prediction may be expressed as the ratio:

$$\frac{\text{Array variance}}{\text{Marginal variance}} = \frac{p(p+2m)(x+p)}{(p+m)^3}.$$

In both limiting processes, i.e. $p \to 0$ and $m \to \infty$, this ratio will tend to zero.

Going back to the array distribution (21) and utilizing equations (10) and (11) we find for the shape coefficients of the array distributions after substitution of (20*a*) and (20*b*)

$$\beta_{1.\,\text{Arr.}} = \frac{(p+3m)^2}{m(p+2m)(x+p)}, \tag{28}$$

$$\beta_{2.\,\text{Arr.}} = 3+\frac{6}{x+p}+\frac{(p+m)^2}{m(p+2m)(x+p)}. \tag{29}$$

For x large and m and p finite we find $\beta_{1.\,\text{Arr.}} \doteqdot 0$, $\beta_{2.\,\text{Arr.}} \doteqdot 3$, which means that irrespective of the magnitudes of the parameters of the bivariate negative binomial the array distributions will tend towards normality for large x. This is well in accordance with observations where we usually have the first array distributions very skew bell- or J-shaped, passing into skew bell-shaped for moderate x and finally approaching normality for the larger x.

For $p \to \infty$ equations (28) and (29) become

$$\beta_{1.\,\text{Arr.}} \doteqdot \frac{1}{m}, \quad \beta_{2.\,\text{Arr.}} \doteqdot 3+\frac{1}{m}.$$

These are the shape coefficients of the Poisson distribution, and as the x's have dropped out we conclude that the shapes of all array distributions are alike. In such a case our bivariate negative binomial has deteriorated into an uncorrelated bivariate Poisson (with $\rho = 0$); hence this is the case of non-proneness.

From the above we may derive a quick but rough rule of thumb method for the detection of proneness from a scattergram. If there is a shape transformation of observed array distributions from the lower x's to the higher ones we have proneness. If not, then we deal with a phenomenon generated by chance alone without proneness as a causal factor.

From the foregoing it is clear that the product-moment correlation ρ of the bivariate negative binomial cannot be interpreted in the same way as a correlation coefficient of the same magnitude derived from a normal bivariate distribution. However, it is still possible to make some useful predictions based on the bivariate negative binomial. In a certain class of selection problems it is required to forecast on the strength of an observed value x whether an individual will belong to one of two classes on the variable y. In such a situation it is more meaningful to work with 'operating characteristics' instead of a single correlation coefficient (Sichel, 1952; Arbous, 1953). Furthermore, without altering their meaning, the principles of operating characteristics may be applied to all bivariate distributions irrespective of their analytical form.

For our special purpose the operating characteristic of absences is defined as the probability of a person with x absences in the predictor period having β or more absences in the follow-up interval. This may be stated as

$$\text{prob.}\,(y \geqslant \beta|x) = 1-\text{prob.}\,(y<\beta|x),$$

$$\text{prob.}\,(y \geqslant \beta|x) = 1-\left(1-\frac{m}{p+2m}\right)^{x+p}\frac{1}{\Gamma(x+p)}\sum_{y=0}^{\beta-1}\frac{\Gamma(y+x+p)}{\Gamma(y+1)}\left(\frac{m}{p+2m}\right)^{y}. \tag{30}$$

3. Testing the admissibility of the working model

The existence or non-existence of proneness in an observed group of individuals cannot be established sufficiently by simply fitting univariate distributions to the observations. The authors agree with Maritz (1950) that correlational (bivariate) analysis is much more

meaningful than ordinary curve fitting. Maritz's point of view has recently been queried by Blum & Mintz (1951). However, it appears to us that univariate analysis sacrifices information contained in the original data. If an individual has a total of z absences in an observational period made up of x occurrences in the first and y in the second half, curve fitting will make use of z only, whereas the correlation approach utilizes the information contained in both x and y. In other words, to test the hypothesis of proneness we anticipate that the correlation approach will be statistically more powerful (in the Neyman-Pearson sense) than the univariate method. As a practical test for the establishment of proneness we therefore recommend:

Split the total observational period into two halves and correlate by product-moment method the absences of individuals in the first period with their respective absences in the second. If we do not know the underlying bivariate distribution which generates the correlation coefficient we have no means of testing for significance. For this reason we should look at low coefficients as rather doubtful in establishing proneness. Even if we have found a large coefficient we cannot do much with it when it comes to the down-to-earth question of predicting absences in the future. For the solution of this problem we must have a model of the bivariate distribution of absences in two equal non-overlapping periods.

If our initial assumptions are correct, then the absences in a single observational period should follow the univariate negative binomial and, therefore, the absences in two equilong, non-overlapping periods may be described by the corresponding symmetrical bivariate distribution.

Scores of investigators have pointed out before us that the univariate negative binomial distribution may be generated by entirely different, often opposing assumptions. A satisfactory fit of observed distributions to theoretical negative binomials, as evidenced by the χ^2-test, means little if we must interpret our data physically.

The Greenwood-Yule model, which we have adapted to absenteeism in this paper, rests on the following premises:

(1) The number of absences incurred by an individual in repeated exposure periods all of equal length t follow a Poisson distribution with mean λ. (Our equation (1) if $k = 1$.)

(2) The mean parameter λ is invariant in time.

(3) The individual λ's in a group of persons observed for one period of length t follow a Pearson Type III law. (Our equation (2).)

Over shorter exposure periods (1) and (2) will often be met. For longer durations (2) is the most exacting of the required conditions.

We may write

$$\lambda = \lambda_1 + \lambda_2,$$

where λ_1 = mean number of absences due to environmental factors, λ_2 = mean number of absences due to individual proneness, λ = total mean number of absences. It is clear that λ will change in time if an individual's environment is altered, i.e. in the event of promotion, better working conditions, different type of work, etc. For this reason one tries to keep environmental conditions as homogeneous as is possible in a modern industrial set-up when studying proneness in a group of workers. Even λ_2, which is person-centred, may vary, as human beings have the habit of learning and adapting themselves if exposed to one and the same situation for any length of time.

In practice it will be found that the χ^2-test is not powerful enough to detect time variations in the mean parameter λ. Due to the small number of observations in intra-personal

investigations (usually of the order 50) the Poisson hypothesis will rarely be rejected. The same argument applies to the Poisson index of dispersion.

Assumption (3) is usually met because a Pearson Type III law is extraordinarily flexible and can represent J-shaped, skew bell-shaped and normal distributions.

As already mentioned we do not expect conditions (1) and (2) to hold exactly in a real life situation. If we assume that the departure from the Poisson hypothesis for individuals is not of such a nature as to falsify completely the assumptions in a given time period, we may proceed to work with our model (and it is for this reason that we prefer to call it a *working* model).

We shall fit a symmetrical bivariate negative binomial to our data. A χ^2-test on the bivariate model, a graphical test for linearity of regression and a comparison between the actual product-moment correlation r and $\hat{\rho}$ as calculated from the estimates of the parameters of the univariate negative binomial (total exposure period) will suffice to indicate the plausibility of the working model.

In Table 1* the absences of 248 shift workers from the same division of a large industrial plant have been recorded for a two-yearly period (1947–8). The product-moment correlation is

$$r = 0{\cdot}728,$$

which indicates absence proneness in the group.

In fitting a symmetrical bivariate negative binomial (equation (13)) to the observed absences of Table 1 we must estimate parameters m and p, where m is the mean of the marginal distributions. This can be done by estimating p and $M = 2m$ of the univariate negative binomial distribution for the total period of exposure.

Formulae, graphs and tables for the estimation of the parameters of a negative binomial distribution were previously given both for the method of moments and the maximum-likelihood solution (Sichel, 1951). Entering the efficiency graphs of this paper with our moment estimates

$$\overline{M} = 9{\cdot}2 \quad \text{and} \quad \overline{p} = 1{\cdot}52,$$

we find

$$\text{Eff.}\,(\overline{p}) \doteqdot 0{\cdot}6,$$

hence we proceed to the maximum-likelihood estimation which yields

$$\hat{m} = \tfrac{1}{2}\hat{M} = 4{\cdot}59072, \quad \hat{p}_{1947/48} = 1{\cdot}58542.$$

Substitution of these values into (13), and evaluation of the theoretical cell frequencies, leads to the expected figures shown in parentheses in Table 1. A χ^2-test comparing observed with expected frequencies of the symmetrical bivariate negative binomial resulted, for 13 degrees of freedom and $\chi^2 = 17{\cdot}0$, in $P = 0{\cdot}20$. On this basis we are unable to reject the hypothesis.

Using the maximum-likelihood estimates we find from equation (17)

$$\hat{\rho} = 0{\cdot}743,$$

which compares satisfactorily with the previously found product-moment correlation

$$r = 0{\cdot}728.$$

We should, however, point out that we have repeatedly observed r to be slightly smaller than $\hat{\rho}$ if the exposure periods are fairly long. Similar observations were made in the field of accidents by Newbold (1927). The reason for this is not quite clear at present.

Substitution of $\hat{\rho}$ and $\hat{p}$ into equation (22) gives the regression line

$$\tilde{y} = 0{\cdot}743x + 1{\cdot}178.$$

* This is a folding table facing p. 90.

Although the observed means follow a linear regression, as required by our model, closer scrutiny of the graph reveals that out of eighteen array means shown, twelve lie below the calculated regression line and only six above. Therefore we have a tendency for the observed number of absences in 1948 to be slightly less than predicted on the basis of the observed absences in 1947. This fact is also borne out by the observed means of our marginal distributions. We have

$$1947: \quad \bar{x} = 4{\cdot}70564 \quad (n = 248),$$

$$1948: \quad \bar{y} = 4{\cdot}47581 \quad (n = 248).$$

For large samples such as we are dealing with, a statistical test of the difference between correlated means of negative binomial distributions is available. When applied to our data it was found that the hypothesis of chance fluctuation in the observed means could not be rejected.

4. THE APPLICATION OF THE WORKING MODEL

If we consider β or more absences per exposure period as worse than tolerable, we certainly shall wish to single out all men who are likely to have β or more absences in any future period. We cannot do this with certainty, even for correlations approaching unity, as proved in the mathematical deductions. On the other hand, the operating characteristic (equation (30)) will tell us with what probability we are able to spot future absence offenders, given their absence numbers in a previous period.

In the preceding section we have treated the 1947 absences as the 'predictor' and the 1948 data as the 'criterion'. However, our 'validation' was done *a posteriori*. In fact, the estimates for m and p were based on the combined period 1947–8. We shall now demonstrate that we may forecast without an initial follow-up by assuming no knowledge of the events in 1948 whatsoever. After having made our predictions we shall compare them with the actual happenings of that year. This then will be a true *a priori* validation.

The standard unit of observational time was defined as one 4-weekly period, so that one calendar year was made up of thirteen 4-weekly periods which included an annual holiday of 3 weeks. For correlational purposes we split the first twelve 4-weekly periods during 1947 into equal halves. The calculated product-moment correlation was $r = 0{\cdot}55$ for $n = 318$. From this we infer absence-proneness in the group. The data conform with the linear regression hypothesis of the bivariate negative binomial. Next we estimated the parameters of the univariate negative binomial for the total year 1947 which, taking cognisance of the 3-weekly annual leave, corresponded to an actual exposure period of 12·25 4-weekly periods. Making use of the tables for the maximum-likelihood solution and the formulae for the standard errors (Sichel, 1951) we arrived at the following estimates of the parameters:

$$\hat{m} = 4{\cdot}9340 \pm 0{\cdot}2452 \quad (n = 318),$$

$$\hat{p} = 1{\cdot}7164 \pm 0{\cdot}1925 \quad (n = 318).$$

A comparison was made between the observed and expected frequencies for 1947 using the above estimates in equation (7). Agreement was satisfactory and, on the strength of the χ^2-test result,

$$P(11{\cdot}419;\ 11) \doteq 0{\cdot}42,$$

we cannot reject the hypothesis of our model.

If we assume that the subjects spread their annual leave evenly over the total year, the average exposure for a 24-weekly period is only 22·615, or measured in 4-weekly units 5·654. Our estimate of m for an actual exposure of 12·25 units was 4·9340, hence we estimate the mean number of absences per 4-weekly exposure to be

$$4{\cdot}9340/12{\cdot}25 = 0{\cdot}40278,$$

in view of equation (8). For 5·654 4-weekly periods we find

$$\hat{m}' = 0{\cdot}40278 \times 5{\cdot}654 = 2{\cdot}277,$$

which compares satisfactorily with the observed means of 2·267 and 2·333. Our estimate of the product-moment correlation based on the parameters of the univariate negative binomial becomes (from equation (17))

$$\hat{\rho} = 2{\cdot}277/(1{\cdot}716 + 2{\cdot}277) = 0{\cdot}57,$$

which compares favourably with $r = 0{\cdot}55$ by ordinary product-moment correlation.

Distributions were also observed for the following exposure periods in 1947:

3 4-weekly periods,
6 4-weekly periods,
8·25 4-weekly periods.

From equations (8) and (9) we computed the expected means and standard deviations by substituting

$$\hat{m} = 0{\cdot}40278, \quad \hat{p} = 1{\cdot}7164,$$

being the estimates for a 4-weekly exposure which in turn were obtained from the year distribution. The observed and expected statistics are given in Table 2. Again the agreement between theory and observation is plausible.

Table 2. *Showing observed and expected statistics for various lengths of exposure*

Number of 4-weekly periods k	$\hat{p}$		$\bar{x}$		s	
	Observed	Expected	Observed	Expected	Observed	Expected
3·00	2·01	1·72	1·12	1·21	1·32	1·43
6·00	1·27	1·72	2·50	2·42	2·72	2·41
8·25	1·93	1·72	3·33	3·32	3·01	3·12
12·25	1·72	1·72	4·93	4·93	4·37	4·37

Before proceeding to the prediction of absences in 1948 we shall derive the standard error of

$$\hat{\rho} = \hat{m}/(\hat{p} + \hat{m}).$$

For large sample sizes we have

$$\operatorname{var}(\hat{\rho}) = \left(\frac{\partial\hat{\rho}}{\partial\hat{p}}\right)^2 \operatorname{var}(\hat{p}) + \left(\frac{\partial\hat{\rho}}{\partial\hat{m}}\right)^2 \operatorname{var}(\hat{m}) + 2\left(\frac{\partial\hat{\rho}}{\partial\hat{p}}\right)\left(\frac{\partial\hat{\rho}}{\partial\hat{m}}\right)\operatorname{cov}(\hat{p}, \hat{m}).$$

After substitution of the appropriate partial derivatives and the variances and covariance given previously by Sichel (1951), we find after replacing the statistics by their expectations:

$$\text{S.E.}(\hat{\rho}) = (1-\rho)\sqrt{\left(\frac{1}{n}\frac{\rho}{p}\left\{1 + \frac{\rho}{p}\left[\sum_{x=2}^{\infty}\frac{\rho^x}{x}B(x,p)\right]^{-1}\right\}\right)}. \tag{31}$$

On the basis of the estimates $\hat{m}$ and $\hat{p}$ derived from the 1947 absence distribution we can *predict* the correlation between the 1947 and 1948 absences, i.e.

$$\hat{\rho}_{47/48} = 4{\cdot}9340/(1{\cdot}7164 + 4{\cdot}9340) = 0{\cdot}742.$$

From (31) we compute the standard error of $\hat{\rho}$ and finally we have

$$\hat{\rho}_{47/48} = 0{\cdot}742 \pm 0{\cdot}024 \quad (n = 318).$$

Of the 318 workers exposed in 1947 only 248 survived employment in 1948. The bivariate distribution for their absences in 1947–8 has already been given in Table 1 and the corresponding product-moment correlation was

$$r_{47/48} = 0{\cdot}728 \quad (n = 248),$$

which checks the predicted correlation within one standard error.

We have not lost sight of the fact that the shrinkage in observed correlation may have been due to the more absence-prone persons leaving employment before a lengthy period of exposure has elapsed. Without giving details we have sufficient evidence to suggest that this is not so.

Substitution of $\hat{m}$ and $\hat{p}$ based on the yearly period of 1947 into equation (30), and making $\beta = 6$ (for the sake of example) gives the operating characteristic

$$\text{prob.}\,(y \geqslant 6|x) = 1 - (0{\cdot}5741)^{x+1{\cdot}7164} \frac{1}{\Gamma(x+1{\cdot}7164)} \sum_{y=0}^{5} \frac{\Gamma(y+x+1{\cdot}7164)}{\Gamma(y+1)} (0{\cdot}4259)^{y}.$$

The predicted O-C curve and the proportion of workers who actually had 6 or more absences in 1948 for each group of individuals with x absences in 1947 have been charted in Fig. 1. Although all observations fall well within the probability belt of individual proportions, we find the majority of points to plot slightly below the theoretical curve. This again is but another manifestation of shrinkage in the mean for 1948.

The O-C curve was based on the absences of 318 subjects in 1947. During the validation year (1948) 70 individuals left their employment so that only 248 could be followed up. To test whether 'restriction' had taken place we calculated the means and standard deviations for the 1947 absences of the two groups of $n_1 = 318$ and $n_2 = 248$. We found

1947 *absences*

$$n_1 = 318 \quad n_2 = 248$$
$$\bar{x}_1 = 4{\cdot}93 \quad \bar{x}_2 = 4{\cdot}70$$
$$s_1 = 4{\cdot}37 \quad s_2 = 4{\cdot}32.$$

The above figures suggest that the slight discrepancies between observations and theory are not due to 'restriction of range'.

In the practical situation we may be interested to single out for curative measures all those who are likely to have 6 or more absences in 1948. By necessity the 'selection' has to be undertaken at the end of 1947. There will always be some few individuals who have say 0, 1, 2 absences in the predictor year and 6 or more in the following period. However, the probability of spotting these is small (see O-C curve, Fig. 1) and we shall do more harm by singling out many for treatment who do not need it than finding the few who will have a bad record in the future period. Clearly what we need is an assurance that the error and extra effort of attending to people who do not need treatment are small in comparison to the gains

originating from a reduction of absenteeism among persons whom we can justifiably call absence-offenders. It appears logical to set the predictor cutting-score α at the same level as β, because if 6 or more absences in 1948 are considered worse than tolerable then the same argument ought to apply to 1947 as well. From the O-C curve we find for $\alpha = 6$ the probability of treating the correct person to be *at least* 0·475.

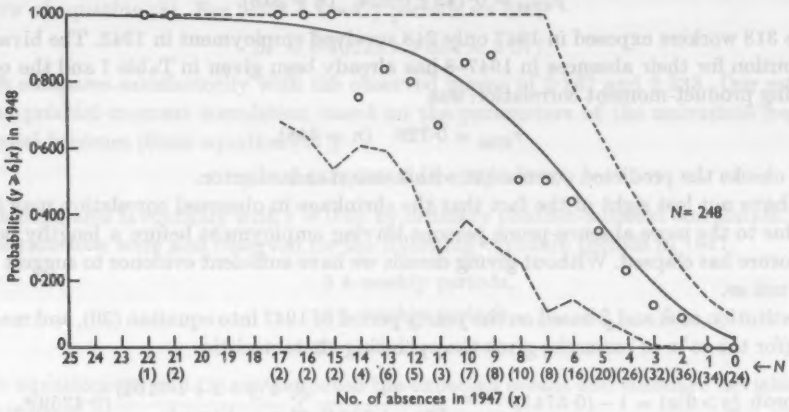


Fig. 1. Operating characteristic with probability limits for observed proportions based on univariate negative binomial distribution for 1947.

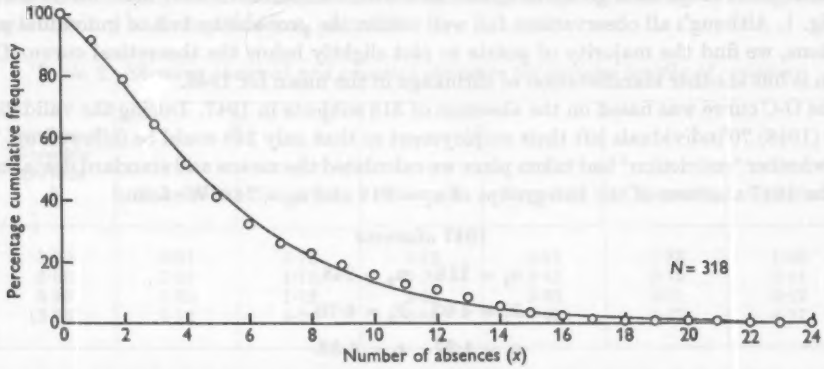


Fig. 2. Cumulative frequency distribution of observed and expected absences in 1947.

In Fig. 2 we have drawn the theoretical and observed cumulative frequency distribution of absences in 1947. From it we may infer what percentage of the working population must be singled out for closer attention with respect to absenteeism for any given level of α.

For $\alpha = 6$ we find that on the average 35 % merit attention. In our particular case we should have selected 103 (32·4 %) from the total group of 318 individuals.

It is unavoidable that some of the potential absentees will terminate employment before the follow-up period has elapsed. Of 103 subjects having had 6 or more absences in 1947, 76 were still employed at the end of 1948. In Table 3 predictions and follow-up results are

compared. Of the 76 absence-offenders in 1947 who were selected for 'treatment' 55 (= 72%) were expected to be absence-offenders in 1948 as well. We actually observed 52 (= 68%) with 6 or more absences in 1948.

Table 3. *Showing comparison between expected and observed numbers of absence-offenders in* 1948

Number of absences in 1947 x	Probability of having 6 or more absences in 1948 $p(x)$	Number of persons exposed in 1947-8 $E(x)$	Number of persons predicted to have 6 or more absences in 1948 $E(x)\,p(x)$	Number of persons who actually had 6 or more absences in 1948 $n(x)$
6	0·475	16	8	7
7	0·568	8	4	4
8	0·652	10	6	5
9	0·726	8	6	7
10	0·786	7	6	6
11	0·840	3	2	2
12	0·881	5	4	4
13	0·911	6	6	5
14	0·931	4	4	3
15	0·948	2	2	2
16	0·961	2	2	2
17	0·971	2	2	2
18	0·979	—	—	—
19	0·985	—	—	—
20	0·990	—	—	—
21	0·994	2	2	2
22	0·997	1	1	1
Totals	—	76	55	52

Bearing in mind that all our predictions were based on the 1947 data alone and that our working model does not take account of factors which may cause significant time variations in individual λ's, we conclude that the agreement between predicted and actual events as revealed by Table III is highly satisfactory.

SUMMARY

In this study a model for absence-proneness has been developed, and we have illustrated the practical use to which the concept of proneness can be put. This application to the needs of industry can be summarized as follows:

(*a*) It is necessary to accumulate one year's data only before preventive action can be taken in respect of absenteeism.

(*b*) These data can be split into two halves for the purpose of establishing the phenomenon of proneness by correlation techniques.

(*c*) In terms of the estimates of the parameters of the distribution observed, a mathematical estimate can be made as to the magnitude of this correlation coefficient if that year's data were compared with the next, which are not yet to hand, and which it is desired to predict.

(*d*) It has been established in our case that, though this mathematical model, strictly speaking, may not be the correct one, it will be sufficiently accurate for practical purposes, provided the two consecutive exposure periods do not exceed *one* year. Furthermore, in any given year the accuracy of theoretical estimates can be checked as has already been done in this study. The investigator is not compelled therefore merely to place implicit faith in the results of the present research. The efficiency with which we can predict the absences of individuals in the second year before these have actually occurred, can then be established in terms of operating characteristics.

(*e*) These operating characteristics will give management information which will enable it to decide on the appropriate remedial measures to be taken in the given circumstances, with a clear indication of the consequences and effectiveness of its actions.

(*f*) When, in actual practice, remedial measures are applied at the end of the first year and before the second year's data have become available, it will not be possible to test the validity of the working model developed here. This does not mean that the theory has broken down entirely. In fact, the efficacy of the measures taken can be judged by the extent to which observation deviates, from theory, over and above the adjustment process suggested above. Where none occurs, this would indicate that the measures had not been effective. Deviations should naturally be for the better. One would expect the observed points to shift uniformly to one side of the curve. It might well be possible in the future to establish along these lines a means of assessing the efficiency of the remedial measures taken.

REFERENCES

Arbous, A. G. (1953). *Tables for Aptitude Testers.* National Institute for Personnel Research (C.S.I.R.) South Africa.

Arbous, A. G. & Kerrich, J. E. (1951). Accident statistics and the concept of accident-proneness. *Biometrics*, **7**, 340.

Blum, M. L. & Mintz, A. (1951). Correlation versus curve fitting in research on accident-proneness. Reply to Maritz. *Psychol. Bull.* **48**, 413.

Fox, J. B. & Scott, J. F. (1943). *Absenteeism: Managements' Problem.* Graduate School of Business Administration Harvard University, No. 29.

Greenwood, M. & Yule, G. U. (1920). An inquiry into the nature of frequency distributions of multiple happenings, etc. *J. R. Statist. Soc.* A **83**, 255.

Lundberg, O. (1940). *On Random Processes and their Application to Sickness and Accident Statistics.* Uppsala: Almquist and Wiksell.

Maritz, J. S. (1950). On the validity of inferences drawn from the fitting of Poisson and negative binomial distributions to observed accident data. *Psychol. Bull.* **47**, 434.

Newbold, E. M. (1927). Practical applications of the statistics of repeated events, particularly to industrial accidents. *J. R. Statist. Soc.* A **90**, 487.

Norris, V. (1951). Health of the schoolchild. *Brit. J. Soc. Med.* **5**, 145.

Russell, D. T., Whitwell, G. P. B. & Ryle, J. A. (1947). Studies in occupational morbidity. (1). *Brit. J. Industr. Med.* **4**, 56.

Sichel, H. S. (1951). The estimation of the parameters of a negative binomial distribution with special reference to psychological data. *Psychometrika*, **16**, 107.

Sichel, H. S. (1952). The selective efficiency of a test battery. *Psychometrika*, **17**, 1.

Sutherland, I. & Whitwell, G. P. B. (1948). Studies in occupational morbidity. (2). *Brit. J. Industr. Med.* **5**, 77.

Table 1. *Observed and expected frequencies during 1947–8 f*

1947 (x-axis) \ 1948 (y-axis)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	7 (11·90)	7 (8·04)	8 (4·43)	1 (2·26)	1 (1·11)	(0·53)	(0·25)	(0·11)	(0·05)	(0·02)	(0·01)	(0·01)	—	—	—	—	—
1	9 (8·04)	6 (8·86)	8 (6·78)	7 (4·42)	1 (2·63)	3 (1·48)	0·80	0·42	0·21	0·11	0·05	0·02	0·01	0·01	—	—	—
2	9 (4·43)	9 (6·78)	6 (6·63)	5 (5·26)	1 (3·69)	3 (2·39)	(1·46)	2 (0·85)	(0·48)	1 (0·26)	(0·14)	(0·07)	(0·4)	(0·02)	(0·01)	(0·01)	—
3	2 (2·26)	8 (4·42)	9 (5·26)	2 (4·92)	4 (3·98)	3 (2·91)	(1·98)	(1·28)	3 (0·79)	1 (0·47)	(0·27)	(0·15)	(0·09)	(0·05)	(0·03)	(0·01)	(0·01)
4	1 (1·11)	7 (2·63)	5 (3·69)	4 (3·98)	3 (3·64)	(2·98)	2 (2·24)	1 (1·58)	1 (1·06)	(0·68)	1 (0·42)	(0·25)	(0·15)	1 (0·09)	(0·05)	(0·03)	(0·01)
5	2 (0·53)	1 (1·48)	1 (2·39)	2 (2·91)	5 (2·98)	2 (2·69)	1 (2·21)	1 (1·69)	(1·23)	1 (0·85)	3 (0·56)	(0·36)	(0·23)	(0·14)	(0·08)	1 (0·05)	(0·03)
6	(0·25)	1 (0·80)	1 (1·46)	2 (1·98)	2 (2·24)	3 (2·21)	1 (1·98)	2 (1·64)	(1·27)	2 (0·94)	(0·66)	1 (0·45)	(0·30)	1 (0·19)	(0·12)	(0·07)	(0·04)
7	(0·11)	(0·42)	(0·85)	1 (1·28)	1 (1·58)	2 (1·69)	1 (1·64)	(1·46)	3 (1·21)	(0·95)	(0·71)	(0·51)	(0·36)	(0·24)	(0·16)	(0·10)	(0·06)
8	(0·05)	(0·21)	1 (0·48)	(0·79)	2 (1·06)	2 (1·23)	(1·27)	(1·21)	2 (1·07)	(0·89)	1 (0·71)	1 (0·54)	(0·39)	(0·28)	(0·19)	(0·13)	(0·08)
9	(0·02)	(0·11)	(0·26)	(0·47)	1 (0·68)	(0·85)	3 (0·94)	(0·95)	2 (0·89)	(0·78)	(0·66)	(0·52)	2 (0·40)	(0·30)	(0·21)	(0·15)	(0·10)
10	(0·01)	(0·05)	(0·14)	1 (0·27)	(0·42)	(0·56)	1 (0·66)	(0·71)	1 (0·71)	(0·66)	(0·57)	1 (0·48)	2 (0·39)	(0·30)	(0·22)	1 (0·16)	(0·12)
11	(0·01)	(0·02)	(0·07)	(0·15)	1 (0·25)	(0·36)	(0·45)	(0·51)	(0·54)	(0·52)	(0·48)	(0·43)	1 (0·35)	(0·29)	(0·22)	(0·17)	(0·13)
12	—	(0·01)	(0·04)	(0·09)	(0·15)	1 (0·23)	(0·30)	(0·36)	1 (0·39)	1 (0·40)	1 (0·39)	(0·35)	(0·32)	(0·26)	1 (0·21,	(0·16)	(0·12)
13	—	(0·01)	(0·02)	(0·05)	(0·09)	1 (0·14)	(0·19)	(0·24)	(0·28)	1 (0·30)	1 (0·30)	1 (0·29)	1 (0·26)	1 (0·23)	(0·19)	(0·15)	(0·12)
14	—	—	(0·01)	(0·03)	1 (0·05)	(0·08)	(0·12)	1 (0·16)	(0·19)	(0·21)	(0·22)	(0·22)	(0·21)	1 (0·19)	(0·19)	(0·14)	(0·11)
15	—	—	(0·01)	(0·01)	(0·03)	(0·05)	(0·07)	1 (0·10)	(0·13)	(0·15)	(0·16)	(0·17)	(0·16)	(0·15)	1 (0·14)	(0·13)	(0·10)
16	—	—	—	(0·01)	(0·01)	(0·03)	(0·04)	(0·06)	(0·08)	(0·10)	1 (0·12)	(0·13)	(0·12)	(0·12)	(0·11)	(0·10)	1 (0·10)
17	—	—	—	(0·01)	(0·01)	(0·01)	(0·03)	(0·04)	(0·05)	(0·07)	1 (0·08)	(0·09)	(0·09)	(0·09)	1 (0·09)	(0·08)	(0·08)
18	—	—	—	—	(0·01)	(0·01)	(0·02)	(0·02)	(0·03)	(0·04)	(0·05)	(0·06)	(0·07)	(0·07)	(0·07)	(0·07)	(0·06)
19	—	—	—	—	—	(0·01)	(0·01)	(0·01)	(0·02)	(0·03)	(0·04)	(0·04)	(0·05)	(0·05)	(0·05)	(0·05)	(0·05)
20	—	—	—	—	—	—	(0·01)	(0·01)	(0·01)	(0·02)	(0·02)	(0·03)	(0·03)	(0·03)	(0·04)	(0·04)	(0·04)
21	—	—	—	—	—	—	—	1 (0·01)	(0·01)	(0·01)	(0·02)	(0·02)	(0·02)	(0·02)	1 (0·03)	(0·03)	(0·03)
22	—	—	—	—	—	—	—	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·02)	(0·02)	(0·02)	(0·02)
23	—	—	—	—	—	—	—	—	—	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·02)	(0·02)
24	—	—	—	—	—	—	—	—	—	—	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)
25	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	(0·01)	(0·01)
26	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
27+	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	30 (28·72)	39 (33·84)	39 (32·52)	25 (28·89)	23 (24·61)	20 (20·44)	9 (16·67)	9 (13·43)	13 (10·71)	7 (8·48)	9 (6·67)	4 (5·22)	6 (4·07)	4 (3·16)	4 (2·45)	2 (1·89)	1 (1·45)

ncies during **1947–8** *for* **248** *shift workers*

8 (y-axis)

3	14	15	16	17	18	19	20	21	22	23	24	25	26	27+	
–	—	—	—	—	—	—	—	—	—	—	—	—	—	—	24 (28·72)
01	—	—	—	—	—	—	—	—	—	—	—	—	—	—	34 (33·84)
02)	(0·01)	(0·01)	—	—	—	—	—	—	—	—	—	—	—	—	36 (32·52)
05)	(0·03)	(0·01)	(0·01)	(0·01)	—	—	—	—	—	—	—	—	—	—	32 (28·89)
09)	(0·05)	(0·03)	(0·01)	(0·01)	(0·01)	—	—	—	—	—	—	—	—	—	26 (24·61)
14)	(0·08)	1 (0·05)	(0·03)	(0·01)	(0·01)	(0·01)	—	—	—	—	—	—	—	—	20 (20·44)
19)	(0·12)	(0·07)	(0·04)	(0·03)	(0·02)	(0·01)	(0·01)	—	—	—	—	—	—	—	16 (16·67)
24)	(0·16)	(0·10)	(0·06)	(0·04)	(0·02)	(0·01)	(0·01)	(0·01)	(0·01)	—	—	—	—	—	8 (13·43)
28)	(0·19)	(0·13)	(0·08)	(0·05)	(0·03)	1 (0·02)	(0·01)	(0·01)	(0·01)	—	—	—	—	—	10 (10·71)
30)	(0·21)	(0·15)	(0·10)	(0·07)	(0·04)	(0·03)	(0·02)	(0·01)	(0·01)	(0·01)	—	—	—	—	8 (8·48)
30)	(0·22)	1 (0·16)	(0·12)	(0·08)	(0·05)	(0·04)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)	—	—	—	7 (6·67)
29)	(0·22)	(0·17)	(0·13)	(0·09)	1 (0·06)	(0·04)	(0·03)	(0·02)	(0·01)	(0·01)	(0·01)	—	—	—	3 (5·22)
26)	1 (0·21)	(0·16)	(0·12)	(0·09)	(0·07)	(0·05)	(0·03)	(0·02)	(0·01)	(0·01)	(0·01)	—	—	—	5 (4·07)
23)	(0·19)	(0·15)	(0·12)	(0·09)	(0·07)	(0·05)	(0·03)	(0·02)	(0·02)	(0·01)	(0·01)	—	—	—	6 (3·16)
19)	(0·19)	(0·14)	(0·11)	(0·09)	(0·07)	1 (0·05)	(0·04)	(0·03)	(0·02)	(0·01)	(0·01)	—	—	—	4 (2·45)
15)	1 (0·14)	(0·13)	(0·10)	(0·08)	(0·07)	(0·05)	(0·04)	(0·03)	(0·02)	(0·02)	(0·01)	(0·01)	—	—	2 (1·89)
12)	(0·11)	(0·10)	1 (0·10)	(0·08)	(0·06)	(0·05)	(0·04)	(0·03)	(0·02)	(0·02)	(0·01)	(0·01)	—	—	2 (1·45)
09)	1 (0·09)	(0·08)	(0·08)	(0·06)	(0·05)	(0·04)	(0·04)	(0·03)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)	(0·01)	2 (1·12)
07)	(0·07)	(0·07)	(0·06)	(0·05)	(0·05)	(0·04)	(0·03)	(0·03)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)	(0·01)	0 (0·86)
05)	(0·05)	(0·05)	(0·05)	(0·04)	(0·04)	(0·04)	(0·03)	(0·02)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)	(0·01)	0 (0·66)
03)	(0·04)	(0·04)	(0·04)	(0·04)	(0·03)	(0·03)	(0·03)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	0 (0·50)
02)	1 (0·03)	(0·03)	(0·03)	(0·03)	(0·03)	(0·02)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	2 (0·38)
02)	(0·02)	(0·02)	(0·02)	(0·02)	(0·02)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	1 (0·01)	(0·01)	1 (0·29)
01)	(0·01)	(0·02)	(0·02)	(0·02)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	0 (0·24)
01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	0 (0·18)
–	—	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	0 (0·13)
–	—	—	—	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	0 (0·11)
–	—	—	—	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·21)	0 (0·31)
4 16)	4 (2·45)	2 (1·89)	1 (1·45)	0 (1·12)	1 (0·86)	2 (0·66)	0 (0·50)	0 (0·38)	0 (0·29)	0 (0·24)	0 (0·18)	0 (0·13)	1 (0·11)	0 (0·31)	248 (248·00)

ON THE SUPERPOSITION OF RENEWAL PROCESSES

BY D. R. COX AND WALTER L. SMITH

Statistical Laboratory, University of Cambridge

Suppose that there are a number of independent *sources* at each of which *events* occur from time to time. The intervals between successive events at any one source are assumed to be independent random variables all with the same distribution, so that each source constitutes a renewal process of a familiar type. The outputs of the sources are combined into one pooled output. Statistical properties of the pooled output are investigated and the results applied to a problem in neuro-physiology.

1. INTRODUCTION

This work was suggested by problems in neuro-physiology, although there are very probably other applications. Suppose that a number of neurons independently send discrete pulses to a common central nerve cell. We shall investigate the relation between the statistical properties of the sequence of impulses from an individual neuron and the corresponding properties of the combined sequence of pulses at the central cell. A very similar problem arose in the recent study by Fatt & Katz (1952) of spontaneous subthreshold activity at motor nerve endings. They found that at the tips of certain many-branching nerve endings, there are a large number of 'active spots' each giving rise to localized electrical pulses in the muscle fibre with which they make common contact. Here the statistical problem is to infer as much as possible about the individual 'active spots' from observations on the sequence of pulses in the muscle fibre.

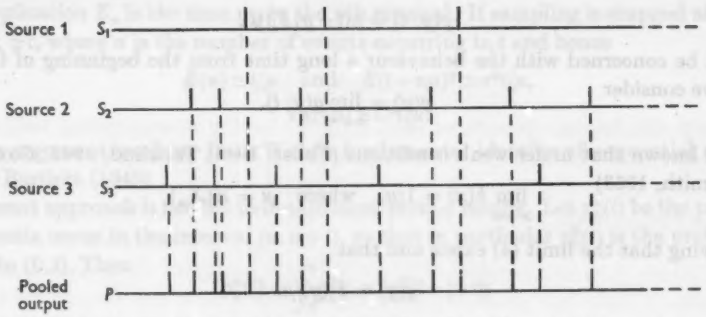


Fig. 1. The pooling of outputs.

We can state the problem formally as follows. Suppose that there are a number of *sources*, at each of which *events* occur from time to time, and that the *outputs* of the sources are combined into one *pooled output*. Fig. 1 illustrates this for three sources. The events are marked by short vertical lines and the outputs of the sources S_1, S_2 and S_3 are combined into one pooled output P. In general, the output of the ith source is defined by an increasing sequence $[t_j^{(i)}]$, where $t_j^{(i)}$ is the time at which the jth event occurs at the ith source. The pooled output is defined by rearranging the $t_j^{(i)}$ into a single increasing sequence. Given the statistical properties of the separate sequences $[t_j^{(i)}]$, we wish to find the statistical properties of the pooled output.

In an earlier paper (Cox & Smith, 1953*a*) we examined what happens when the individual sequences S_i are strictly periodic, i.e. when S_i is $[\theta_i, 2\theta_i, 3\theta_i, \ldots]$. In the present paper we shall suppose that the intervals between successive events on any one source are independent random variables, all with the same distribution, and that random variables associated with different sources are independent. Thus the output of one source forms a *renewal process* of the type that has been extensively studied (Feller, 1941).

In §§ 2, 3 we give some properties of single renewal processes. In § 4 we develop the main properties of the pooled output and in § 5 we investigate the limiting behaviour as the number of sources becomes large. A numerical application to experimental data is described in § 6.

2. The delay distribution for a single source

In this section we consider a single source. Let the intervals between successive events be a sequence $[X_n]$ of independent identically distributed random variables all with an absolutely continuous distribution function $F(x)$, and a frequency function $f(x)$. The nth event occurs at time $Z_n = X_1 + \ldots + X_n$, and if $f_n(x)$ is the frequency function of Z_n,

$$h(x) = \sum_{n=1}^{\infty} f_n(x) \tag{1}$$

is the total density of events at x, irrespectively of serial number.

Fix a time t and define the *delay* $Y(t)$ to be the time measured from t back to the immediately preceding event. The frequency function $g(y; t)$ of $Y(t)$ will be called the delay function at t. Since the total density of events at x is $h(x)$, and the chance that the interval between events exceeds y is

$$F_c(y) = 1 - F(y), \tag{2}$$

we have

$$g(y; t) = h(t-y)\, F_c(y). \tag{3}$$

We shall be concerned with the behaviour a long time from the beginning of the process so that we consider

$$g(y) = \lim_{t \to \infty} g(y; t). \tag{4}$$

It is well known that under weak conditions (Feller, 1941; Täcklind, 1945; Cox & Smith, 1953*b*; Smith, 1953)

$$\lim_{x \to \infty} h(x) = 1/\mu, \quad \text{where} \quad \mu = E(X_n), \tag{5}$$

thus proving that the limit (4) exists and that

$$g(y) = F_c(y)/\mu. \tag{6}$$

We call $g(y)$ the *equilibrium delay function* corresponding to the *parent distribution* $f(x)$.

Conversely, given $g(y)$, $f(x)$ is determined by

$$f(x) = -\frac{1}{y(0)} \frac{dg(x)}{dx}. \tag{7}$$

The derivation of (6), and hence also of (7), depends only on the assumptions that

(i) the distribution function of the intervals between successive events is $F(x)$;

(ii) the density of events at t tends to $1/\mu$ as t tends to infinity.

No use is made of the independence of successive intervals other than in the proof of (ii). We shall later apply (7) when conditions (i) and (ii) can be verified although successive intervals are not independent.

It is intuitively obvious, and can be proved, that the equilibrium frequency function of the time interval measured forward from a fixed point to the next succeeding event is the same as that of the equilibrium delay.*

It is easy to express the moments of the delay in terms of the moments of the parent distribution and, in particular, to show that the mean delay is $(\sigma^2+\mu^2)/2\mu$, where μ and σ are the mean and standard deviation of the parent distribution.

3. The number of events in intervals of given length

This section is concerned with the probability distribution and variance of the number of events falling in intervals of given length t. We shall deal mainly with the equilibrium behaviour for intervals situated a long way from the origin, i.e. with the limiting behaviour for intervals $(u, u+t)$ as u tends to infinity. The principal result is that if $V(t)$ is the variance of the number of events in an interval t,

$$V(t) \sim \frac{C^2 t}{\mu} \quad \text{as} \quad t \to \infty, \tag{8}$$

where C is the coefficient of variation of the parent distribution. Further terms in the expansion of $V(t)$ will be obtained. Feller (1949) proved (8) in his well-known paper on recurrent events; the main differences between his and our work are that Feller considered discrete, not continuous, time, and that our problem is easier since it is concerned with equilibrium behaviour.

A heuristic proof of (8) follows easily from two equations in sequential analysis (Wolfowitz, 1947). If $Z_n = X_1 + \ldots + X_n$ is the cumulative sum when sampling stops after n 'events', we have

$$E(Z_n) = E(n)\,\mu \quad \text{and} \quad E(Z_n - n\mu)^2 = E(n)\,\sigma^2. \tag{9}$$

In our application Z_n is the time up to the nth renewal. If sampling is stopped after a long time t, $Z_n \simeq t$, where n is the number of events occurring in t and hence

$$E(n) \simeq t/\mu \quad \text{and} \quad E(t - n\mu)^2 \simeq \sigma^2 t/\mu,$$

i.e.

$$\text{var}\,(n) \simeq C^2 t/\mu. \tag{10}$$

A similar argument working from Wald's fundamental identity of sequential analysis is given by Bartlett (1949).

A different approach is needed to obtain more precise results. Let $p_k^u(t)$ be the probability that k events occur in the interval $(u, u+t)$, so that in particular $p_k^0(t)$ is the probability of k events in $(0, t)$. Then

$$p_k^u(t) = \int_0^t g(y;\, u)\, p_k^0(t-y)\, dy. \tag{11}$$

By (3) and (5) $g(y; u)$ is, for sufficiently large u, bounded in $0 \leqslant y \leqslant t$; also $0 \leqslant p_{k-1}^0(t) \leqslant 1$. Therefore the theorem of bounded convergence gives that

$$\lim_{u\to\infty} \int_0^t g(y;\, u)\, p_{k-1}^0(t-y)\, dy = \int_0^t g(y)\, p_{k-1}^0(t-y)\, dy,$$

thus proving the existence of a limit for $p_k^u(t)$ as $u \to \infty$. We write

$$p_k^e(t) = \lim_{u\to\infty} p_k^u(t), \tag{12}$$

and call $p_k^e(t)$ the *equilibrium distribution* of events.

* The delay is the same as the residual life-time discussed by Doob (1948), and the delay and the forward interval are the same as two quantities defined by Smoluchowski (Bartlett, 1953).

By (6) and (12),

$$p_k^e(t) = \frac{1}{\mu}\int_0^t \{p_{k-1}^0(\eta) - F(t-\eta)\, p_{k-1}^0(\eta)\}\, d\eta.$$

But

$$\int_0^t p_k^0(\eta)\, d\eta = \int_0^t d\eta \int_0^\eta f(\eta-\xi)\, p_{k-1}^0(\xi)\, d\xi$$

$$= \int_0^t F(t-\eta)\, p_{k-1}^0(\eta)\, d\eta,$$

so that

$$p_k^e(t) = \frac{1}{\mu}\int_0^t \{p_{k-1}^0(\eta) - p_k^0(\eta)\}\, d\eta. \tag{13}$$

If we define moment-generating functions by

$$M^0(\theta, t) = \sum_{k=0}^{\infty} e^{k\theta}\, p_k^0(t), \quad M^e(\theta, t) = \sum_{k=0}^{\infty} e^{k\theta}\, p_k^e(t); \tag{14}$$

(13) leads to

$$M^e(\theta, t) = 1 - \frac{(1-e^\theta)}{\mu}\int_0^t M^0(\theta, \eta)\, d\eta. \tag{15}$$

Let $m_1^u(t)$, $m_2^u(t)$, ... be the moments about zero of the distribution $p_k^u(t)$, where $u = 0, e$. From (15) we deduce that when the appropriate quantities exist

$$m_1^e(t) = t/\mu, \tag{16}$$

$$m_2^e(t) = t/\mu + \frac{2}{\mu}\int_0^t m_1^0(\eta)\, d\eta. \tag{17}$$

If $V^e(t)$ is the equilibrium variance, (16) and (17) give

$$V^e(t) = \frac{1}{\mu}\int_0^t \psi(\eta)\, d\eta, \tag{18}$$

where

$$\psi(t) = 1 + 2m_1^0(t) - 2t/\mu. \tag{19}$$

We denote the Laplace transform of a function by the corresponding bold symbol; for example,

$$\boldsymbol{\psi}(s) = \int_0^\infty e^{-st}\psi(t)\, dt. \tag{20}$$

Then it may be shown from (19) that

$$\boldsymbol{\psi}(s) = \frac{1}{s} + \frac{2\mathbf{f}(s)}{s[1-\mathbf{f}(s)]} - \frac{2}{\mu s^2}. \tag{21}$$

Let $\mu, \mu_2, \ldots$ be the moments about zero of $f(x)$ and let C be the coefficient of variation, $C^2 = (\mu_2 - \mu^2)/\mu^2$. Then expansion of (21) gives, as s tends to zero,

$$\boldsymbol{\psi}(s) \sim \frac{C^2}{s} + \frac{3\mu_2^2 - 2\mu\mu_3}{6\mu^3}. \tag{22}$$

This suggests that as $t \to \infty$

$$V^e(t) = \frac{1}{\mu}\int_0^t \psi(\eta)\, d\eta \sim \frac{C^2 t}{\mu} + \frac{3\mu_2^2 - 2\mu\mu_3}{6\mu^4}; \tag{23}$$

a rigorous proof of (23) may be derived, under general conditions, from results in renewal theory; see, for example, Owen (1949) and, more generally, Smith (1953). A similar formula has been indicated by Feller (1948).

Equation (23) shows the behaviour of $V^e(t)$ for large t. We also require the behaviour for small t, which may be obtained as follows. By (18)

$$V^e(t) = \frac{t}{\mu} + 2\int_0^t \left\{ m_1^0(\eta) - \frac{\eta}{\mu} \right\} d\eta, \tag{24}$$

and it is easy to show from this that if as t tends to zero

$$\int_0^t f(x)\,dx = O(t^\beta) \quad (\beta > 0), \tag{25}$$

then

$$V^e(t) = \frac{t}{\mu} + o(t). \tag{26}$$

Thus near $t = 0$ the variance-time curve behaves like that of a completely random series.

If $f(x)$ is given explicitly it may be possible to obtain more precise results than (23) and (26). For example, if $f(x)$ is of the χ^2 type with 2ν degrees of freedom, $C^2 = 1/\nu$ and (23) becomes

$$V^e(t) \sim \frac{t}{\nu\mu} + \frac{1}{6}\left(1 - \frac{1}{\nu^2}\right) \quad \text{as} \quad t \to \infty. \tag{27}$$

$V^e(t)$ may, of course, be found exactly, although the calculation is tedious for large ν. As an example, if $f(x)$ is of the χ^2 type with 6 degrees of freedom and mean 3,

$$\mathbf{f}(s) = \frac{1}{(1+s)^3}$$

and (20) and (21) lead to

$$V^e(t) = \frac{t}{9} + \frac{4}{27}\left(1 - e^{-\frac{3}{2}t}\cos\frac{\sqrt{3}\,t}{2}\right). \tag{28}$$

We may compare this with

$$V^e(t) \sim \frac{t}{9} + \frac{4}{27} \tag{29}$$

obtained from (27).

4. The pooling of outputs

Suppose that the outputs of N sources are pooled in the way described in § 1. We shall assume that the outputs of the individual sources are of the type considered in §§ 2, 3, that different sources are independent, and finally, for convenience, that all the sources have the same parent distribution $F(x)$. We shall consider throughout the equilibrium behaviour a long time after the start of the process.

First we calculate the frequency distribution of the interval between successive events in the pooled output. For each source we define delay random variables, Y_i, where Y_i is the time from a fixed epoch back to the immediately preceding event on the ith source. If Y is the corresponding random variable for the pooled output

$$Y = \min(Y_1, \ldots, Y_N),$$

so that since the N sources are independent,

$$\text{prob}(Y \geqslant y) = \left[\int_y^\infty g(x)\,dx\right]^N = \left[\int_y^\infty \frac{F_c(x)}{\mu}\,dx\right]^N. \tag{30}$$

Therefore the delay function for the pooled output is, on differentiating,

$$\frac{NF_c(y)}{\mu}\left[\int_y^\infty \frac{F_c(x)}{\mu}\,dx\right]^{N-1}.$$

By (7) it follows that the frequency function of the interval between successive events on the pooled output is

$$f_p(y) = -\frac{d}{dy}\left\{F_e(y)\left[\int_y^\infty \frac{F_e(x)}{\mu}\,dx\right]^{N-1}\right\}. \tag{31}$$

For example, if $f(y) = e^{-y}$, (31) gives

$$f_p(y) = N e^{-Ny},$$

expressing the obvious fact that if the separate outputs are completely random so is the pooled output. As another example suppose that the parent distributions are rectangular over $(0, 1)$. Then (31) gives

$$f_p(y) = (2N-1)(1-y)^{2N-2}.$$

As $N \to \infty$,

$$f_p(y) \sim 2N e^{-2Ny}. \tag{32}$$

In § 5 it will be shown that the limiting form is exponential in general.

The equilibrium distribution for the number of events falling in an interval of length t may be found by convoluting the distributions for the individual sources, so that in particular the moment-generating function is

$$[M^e(\theta, t)]^N, \tag{33}$$

and the equilibrium variance of the number of events in an interval is

$$V_p(t) = N V^e(t). \tag{34}$$

Thus

$$V_p(t) \sim \frac{NC^2 t}{\mu} = C^2 t\lambda \quad \text{as} \quad t \to \infty, \tag{35}$$

where $\lambda^{-1} = \mu/N$ is the mean interval between events on the pooled output. Also

$$V_p(t) \sim t\lambda \quad \text{as} \quad t \to 0, \tag{36}$$

and the intercept of the asymptote is, if $f(x)$ is of the χ^2 form,

$$\frac{N}{6}(1 - C^4).$$

Thus if we are given the $V_p(t)$ curve for a pooled output of the type we are considering, C may be estimated by

$$\hat{C}^2 = \text{slope of asymptote} \times \text{mean interval between events}. \tag{37}$$

To estimate N some further assumption is required about the parent distribution $f(x)$. For example, if an estimate is calculated from the intercept of the asymptote, the third moment of $f(x)$ must be known. The simplest procedure is to assume that $f(x)$ is of the χ^2 type, when

$$\hat{N} = \frac{6 \times \text{intercept of asymptote}}{1 - \hat{C}^4} \tag{38}$$

is a suitable estimate. If C is appreciably less than one the estimate is insensitive to changes in the form of $f(x)$.

5. Large number of sources

We now consider briefly the form of the pooled output when the number of sources is large. We assume the weak condition (25) on the parent distribution, namely, that there exists $\beta (0 < \beta < 1)$ such that for small t

$$\int_0^t f(x)\,dx = O(t^\beta). \tag{39}$$

It follows that $$M^0(\theta, t) = 1 + O(t^\beta), \tag{40}$$

so that by (15) $$M^e(\theta, t) = 1 - \frac{(1-e^\theta)}{\mu}\{t + O(t^{1+\beta})\}. \tag{41}$$

Express t as a multiple of the mean interval on the pooled output by putting

$$t = \frac{\tau\mu}{N}.$$

Then we have for the moment-generating function of the pooled output, by (33),

$$\left\{M^e\left(\theta, \frac{\mu\tau}{N}\right)\right\}^N = \left\{1 - (1-e^\theta)\left[\frac{\tau}{N} + O\left(\frac{\tau^{1+\beta}}{N^{1+\beta}}\right)\right]\right\}^N \sim e^{-\tau(1-e^\theta)}. \tag{42}$$

Therefore the number of events falling in an interval of length $\mu\tau/N$ tends to be distributed in a Poisson distribution as N tends to infinity. In particular, the interval between successive events tends to be distributed in an exponential distribution, as can also be proved directly from (31). It does not follow from (42) that pairs of intervals tend to be independent although this is in fact true; we omit the proof.

Thus, for large N the pooled output is 'locally random' in the sense that it has the properties of a completely random series, provided that we consider the behaviour over intervals t small compared with the individual mean recurrence times μ. This result may be expected to hold for a large number of independent sources, irrespective of the precise nature of the output of the single sources. A further particular case has been considered by Cox & Smith (1953*a*). The local approach to complete randomness is of general interest, because there are numerous practical applications of stochastic processes in which complete randomness is customarily assumed in situations where a pooled output is being considered. For instance, the sequence of calls at a telephone exchange is the pooled output of the calls originating from individual subscribers. Hence the overall sequence may be expected to be locally random.

6. Application

We now consider the practical application of our results. Fatt & Katz, in the work referred to above, obtained a sequence of over 800 pulses, being the 'pooled output' of a number of active spots. They showed that the distribution of intervals between successive pulses was fitted satisfactorily by the exponential curve, to be expected for a completely random sequence, although they were careful to point out that the exponential curve was consistent with strictly periodic individual outputs.

We consider three possible structures for the sequence:

(i) the individual sources, and hence also the pooled output, are completely random;

(ii) the sequence is the pooled output of N strictly periodic sources;

(iii) the sequence is the pooled output of N independent sources of the type considered in the present paper.

We wish to decide which possibility holds, in case (ii) to estimate the number, N, of active spots and in case (iii) to estimate both N and the coefficient of variation, C, of intervals on one source.

To do this we use the variance-time curve, $V(t)$. If the sequence is completely random

$$V(t) = t\lambda, \tag{43}$$

where λ^{-1} is the mean interval between successive events in the pooled output. In case (ii), $V(t)$ has been shown in a previous paper to oscillate for large t about an average of $\frac{1}{6}N$. In case (iii) we have shown (35) that

$$V(t) \sim C^2 t\lambda, \quad \text{for large } t, \tag{44}$$

Table 1

t (sec.)	$\hat{V}(t)$	S.E. of $\hat{V}(t)$	$\hat{\lambda}t$
2	11·2	1·41	11·0
4	23·1	3·45	22·0
8	39·9	9·39	44·1
16	54·1	26·8	88·1
32	83·4	79·4	176·2
48	189·7	154	264·3

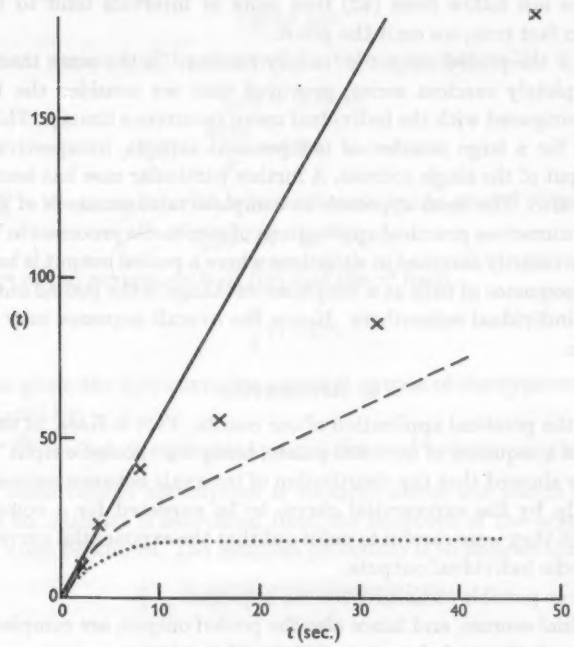


Fig. 2. Variance-time curves. × × × × experimental values. —— theoretical for completely random series. ..... theoretical for 100 strictly periodic sources. – – – – – theoretical for 100 stochastic sources, $C = 1/\sqrt{3}$.

and that C^2 and N can be estimated from the slope and intercept of the asymptote. In all three cases

$$V(t) \sim t\lambda \quad \text{for small } t. \tag{45}$$

The estimation of $V(t)$ from experimental data has been described in the earlier paper (Cox & Smith, 1953*a*). The experimental series is divided into sections of length τ and the

number of events in each section counted. (A suitable choice for τ is approximately one-half the smallest value of t for which $V(t)$ is required.) From the discrete time series so formed, an estimate $\hat{V}(r\tau)$ of $V(r\tau)$ may be calculated by a process of successive addition and squaring, that will not be described in detail here. When the sequence is completely random, $\hat{V}(r\tau)$ is an unbiased estimate of $V(r\tau)$ and its standard error may be calculated.

For Fatt & Katz's data we find the values given in Table 1 and illustrated in Fig. 2. The mean interval between events is $\hat{\lambda}^{-1} = 0{\cdot}220$ sec. The period of observation is 174 sec., so that, for example, $t = 32$ sec. is an appreciable proportion of the series available for analysis. This is why the standard error of $\hat{V}(32)$ is very large. In no case does $\hat{V}(t)$ differ from its value $\hat{\lambda}t$ for a completely random series by much more than its standard error, so that there is no evidence that the series is not completely random.

If the number of sources were large, or the coefficient of variation were near 1, our method would be insensitive. As examples we have drawn on Fig. 2 theoretical curves for

(*a*) 100 strictly periodic sources (approximate curve);

(*b*) 100 stochastic sources with $C = 1/\sqrt{3}$, the intervals on any source being proportional to χ^2 variables with six degrees of freedom.

If the method were to be used extensively it would be worth-while investigating the distribution of $\hat{V}(t)$ for the alternative sequences of types (ii) and (iii), hence developing methods of determining confidence intervals for N and C.

We are extremely grateful to Prof. B. Katz, F.R.S., and Dr P. Fatt, University College, London, for supplying us with the experimental results analysed in § 6, and for comments on the paper.

REFERENCES

Bartlett, M. S. (1949). *J. R. Statist. Soc.* B, **11**, 211.
Bartlett, M. S. (1953). *Proc. Camb. Phil. Soc.* **49**, 263.
Cox, D. R. & Smith, W. L. (1953*a*). *Biometrika*, **40**, 1.
Cox, D. R. & Smith, W. L. (1953*b*). *Skand. Aktuar.* **36**, 139.
Doob, J. L. (1948). *Trans. Amer. Math. Soc.* **63**, 422.
Fatt, P. & Katz, B. (1952). *J. Physiol.* **117**, 109.
Feller, W. (1941). *Ann. Math. Statist.* **12**, 243.
Feller, W. (1948). *Courant Anniversary Volume*, p. 105. New York: Interscience Publishers.
Feller, W. (1949). *Trans. Amer. Math. Soc.* **67**, 98.
Owen, A. R. G. (1949). *Proc. Roy. Soc.* B, **136**, 67.
Smith, W. L. (1953). *Proc. Roy. Soc. Edinb.* **64**, 9.
Täcklind, S. (1945). *Skand. Aktuar.* **28**, 68.
Wolfowitz, J. (1947). *Ann. Math. Statist.* **18**, 215.

CONTINUOUS INSPECTION SCHEMES

By E. S. PAGE

Statistical Laboratory, University of Cambridge

1. Introduction

1·1. *Preliminary remarks*

Whenever observations are taken in order it can happen that the whole set of observations can be divided into subsets, each of which can be regarded as a random sample from a common distribution, each subset corresponding to a different parameter value of the distribution. The problems to be considered in this paper are concerned with the identification of the subsamples and the detection of the changes in the parameter value. Such problems can arise in a number of fields of application. For example, in an experiment in extrasensory perception the proportion of correct answers given by a subject in response to a series of questions may change during the course of the experiment, and it may be desired to estimate the position of the change or to stop the experiment when a change is noticed. Again, in a psychological experiment in which a subject is required to guess the colour of the next ball to be drawn at random with replacement from a bag containing balls of two colours the subject's proportion of correct guesses in a series of trials with the same bag of balls may change as he gains some knowledge of the constitution of the bag from the results of earlier guesses; it may be of interest to estimate the point at which the change took place. More widely known are the occurrences in industry of problems of detecting changes in the quality of the output from a continuous production process. Some such processes maintain an approximately constant quality of output for considerable periods; occasionally, probably because of a fault at some point of the process, the quality worsens and a large proportion of the output becomes unacceptable. The quality of the output may be assessed by some measurable characteristic (e.g. when the length of articles is normally distributed with constant variance the mean length may be used as an indication of quality), or by the fraction of the output that fails to meet given specifications. In general, it will be possible to assign a quality number, θ, to the output which may be taken as a parameter of the distribution. We are interested in the changes in θ.

One of the simplest criteria for detecting a change in the mean, θ, of the distribution is a weighted sum of the last few, k say, observations, i.e. a moving average. If k is small, large changes in θ will be detected rapidly but small changes only slowly; on the other hand, a larger value of k will be required for the best detection of small changes in θ but then large changes will be noticed later owing to the moving average damping the effect of a single extreme observation. In general, the consequences of rules based on moving averages are difficult to evaluate. The theory of one such rule for identifying the subsamples in observations from a binomial population of changing mean has been given by Anscombe, Godwin & Plackett (1947).

It will be convenient in what follows to use the terminology of the industrial applications in view of its greater familiarity.

1·2. *Detection of a change in the parameter*

The first problem to be considered here is that of detecting a change in the parameter θ. It is this problem that process inspection schemes are designed to solve; it is required to detect a deterioration in the quality of the output from a continuous production process. When such a deterioration is suspected some action is taken; for example, the production may be suspended and a machine reset. A widely used scheme consists of examining samples of a fixed size at regular intervals of time; a statistic of the sample (e.g. mean, range, or number of defectives) is plotted on a control chart and corrective action is taken if the point falls outside control limits drawn on the chart (e.g. Dudding & Jennett, 1942; Duncan, 1952; Shewhart, 1931). In most cases of practical interest there is a probability of one that some point will eventually fall outside the limits and action will then be taken even when there is no change in the quality number, θ. If the fraction of the output that is sampled remains constant the amount of output produced before action is taken is proportional to the number of articles inspected. We are led, therefore, to consider what we shall call the *average run length* function for a process inspection scheme.

DEFINITION. When the quality remains constant the *average run length* (A.R.L.) of a process inspection scheme is the expected number of articles sampled before action is taken.

The A.R.L. is clearly a function of θ, the quality number.

When the quality of the output is satisfactory the A.R.L. is a measure of the expense incurred by the scheme when it gives false alarms, i.e. Type I errors (Neyman & Pearson, 1936). On the other hand, for constant poor quality the A.R.L. measures the delay and thus the amount of scrap produced before the rectifying action is taken, i.e. Type II errors. This measure is one of several suggested by Aroian & Levene (1950); another is the probability that action is taken within n observations.

For the control chart scheme described, on constant quality, the number, r, of samples each of N articles examined before action is taken is a geometric variable such that

$$\Pr(r=k) = P^{k-1}(1-P), \quad \text{where} \quad P = P(\theta),$$

the probability that a given sample point falls between the control limits. $\mathscr{E}(r) = (1-P)^{-1}$ and the A.R.L. function is

$$L(\theta) = N/(1-P). \tag{1}$$

Although the results of previous samples are recorded on the chart none is used by the above process inspection rule; in order to avoid some of this loss, warning lines within the control limits are often drawn on the chart and the rule amended to read: 'Take action if any point falls outside the control limits or if l out of a sequence of m points fall outside the warning lines.' The A.R.L. for rules of this type can be calculated by enumerating the possible combinations of the positions of the last $m-1$ points and treating them as the states of a discrete Markov process; the methods described, for example, in Bartlett's paper (1953), will then give the expected number of samples before a combination demanding action is observed.† Such schemes will not be considered further here.

† When a chart is used for controlling the mean of a normal population, statements are sometimes made in the literature such as: '4 out of 16 points falling outside warning limits at $\mu \pm 1{\cdot}96\sigma/\sqrt{N}$ are equivalent to one point outside control limits at $\mu \pm 3{\cdot}09\sigma/\sqrt{N}$.' If the process mean is μ, these two events have the same probability, but rules using one or both of these criteria for action will have different A.R.L. functions and therefore a different effect in practice.

The process inspection rules described so far are based either on a single point recorded on the control chart or on a fixed number of the most recently recorded points. Such rules will therefore fail to make use of all the information that is available on the chart. Again, with these rules the best sample size and the position of the control limits vary with the magnitude of the change in the quality number that it is important to detect (Page, 1954). Consequently a rule that is optimum for a certain magnitude of change will not be optimum for other magnitudes of change, and in some cases the loss in efficiency may be serious. For example, if the best sample size for a single sample scheme to detect a change from θ_0 to θ_1 is N, the A.R.L. of this scheme is not less than N for any value of θ; it may be possible by some other method to detect a large change in fewer than N observations. In the following sections we develop rules that use all the observations since action was last taken and that are suitable for the detection of any magnitude of change in the parameter.

2. One-sided process inspection schemes

2·1. *A transition scheme*

Process inspection schemes that are designed to detect deviations in θ in only one direction will be called *one-sided* schemes. We consider first a simple example of a rule to control the fraction of defective articles produced by an industrial process.

Suppose samples of twenty articles are taken every hour; let d_n be the number of defectives in the nth sample after action was last taken. Consider the rule:

$$\left.\begin{aligned} &\text{Take action if} \quad d_n \geqslant 3, \\ &\text{or} \qquad\quad d_n + d_{n-1} \geqslant 4, \\ &\text{or} \quad d_n + d_{n-1} + d_{n-2} \geqslant 5, \\ &\qquad \text{etc.} \\ &\text{i.e.} \quad \sum_{i=0}^{r} d_{n-i} \geqslant 3 + r \text{ for at least one } r \quad (0 \leqslant r \leqslant n-1). \end{aligned}\right\} \tag{2}$$

After each sample is taken, the total number of defectives in the last one, two, three, etc., samples is examined and action is taken if, over any sequence of samples, the average number of defectives per sample is 'much' more than one. If 5 % defective is the critical quality level for the process, so that any worse quality requires action to be taken, the above rule would be suitable.

With this rule the decision whether or not to take action is made after each sample and all the previous samples are used in making the decision. This rule may therefore be regarded as transitional between those rules for which the decisions are made after each sample on the results of a fixed number of samples and those for which a decision is made after each observation considering all previous observations (the latter type of rule will be called a *sequential process inspection scheme*).

The action criteria (2) may be specified in a different way. Let a score $+19$ be assigned to each defective occurring in a sample, and -1 for each non-defective. Let $S_n = \sum_{i=1}^{n} x_i$, where x_n is the total score in the nth sample after action was last taken. Then (2) is equivalent to:

$$\text{Take action if } S_n - S_{n-r} \geqslant 40 \text{ for any } r \quad (1 \leqslant r \leqslant n), \tag{3}$$

or, what is again equivalent,

$$\text{Take action if } S_n - \min_{0 \leqslant i < n} S_i \geqslant 40. \tag{4}$$

When the cumulative score S_n is plotted on a chart the mean path when the fraction defective, θ, is constant and less than 5 % is below the horizontal. If $\theta > 0{\cdot}05$, the mean path is above the horizontal so that the action criterion (4) can be expected to be satisfied speedily. Accordingly this rule would be appropriate for controlling a process for which more than 5 % defectives could not be tolerated.

In the above scheme when the quality is satisfactory the mean path on the cumulative sum diagram is downwards, and a deterioration in quality is detected by an upward change of direction in the mean path. A similar criterion for controlling the parameter of other distributions when deviations in only one direction are important is provided by the following rule.

Rule 1. Take samples of fixed size N at regular intervals; assign a score x_k to the kth sample and plot the cumulative score $S_n = \sum_{k=1}^{n} x_k$ on a chart:

$$\text{Take action if } S_n - \min_{0 \leqslant i < n} S_i \geqslant h. \tag{5}$$

The A.R.L. for this rule will be obtained as a special case of that for the rule of the next section.

2·2. *A sequential scheme*

In what follows we shall suppose that observations are recorded singly at regular intervals, and that a decision whether or not to take action because of a suspected deterioration in the output is made after each observation. A rule fulfilling these requirements is obtained by modifying Rule 1.

Rule 2. Take observations at regular intervals; assign a score x_k to the kth observation and plot the cumulative score $S_n = \sum_{k=1}^{n} x_k$ on a chart. Take action if (5) is satisfied.

The system of scoring is chosen so that the mean sample path on the chart when quality is satisfactory is downwards, i.e. of negative gradient, and is upwards when quality is unsatisfactory (Fig. 1).

In order to evaluate the A.R.L., define

$$\left.\begin{aligned} S'_n &= \max\,(S'_{n-1} + x_n, 0) \quad (n \geqslant 1) \\ S'_0 &= 0, \end{aligned}\right\} \tag{6}$$

so that $S'_n = 0$ whenever $S_n < \min_{0 \leqslant i < n} S_i$. The condition (5) is then equivalent to:

$$\text{Take action after the } n\text{th observation if } S'_n \geqslant h. \tag{7}$$

It can now be seen that this rule breaks up into a sequence of Wald sequential tests with boundaries at $(0, h)$ and initial score zero.† The test is reapplied when the previous test ends on the lower boundary and action is taken when a test ends on the upper boundary.

Notation. Let the probability that a Wald test with boundaries $(0, h)$ and initial score Z end on the lower boundary be $P(Z)$ and the average sampling numbers, unconditional, conditional upon the test ending on the lower boundary, and conditional upon the test ending on the upper boundary, be $N(Z)$, $N_1(Z)$ and $N_2(Z)$ respectively. A test that ends upon the lower boundary will be called an 'acceptance test'; conversely for a 'rejection test'.

† The procedure in which observations $x_1, x_2, \ldots$ are taken so long as $a < Z_i < b$, where $Z_i = Z_{i-1} + x_i$, $Z_0 = x_0$ will be termed a Wald sequential test with boundaries (a, b) and initial score x_0.

The probability that there are r acceptance tests before a rejection test occurs is $\{P(0)\}^r\{1-P(0)\}$, and the expected number of such tests is therefore

$$\sum_{r=1}^{\infty} r\{P(0)\}^r\{1-P(0)\} = P(0)/\{1-P(0)\}. \tag{8}$$

It follows that the A.R.L. for this rule is

$$L = \frac{P(0)}{1-P(0)} N_1(0) + N_2(0)$$
$$= \frac{N(0)}{1-P(0)}. \tag{9}$$

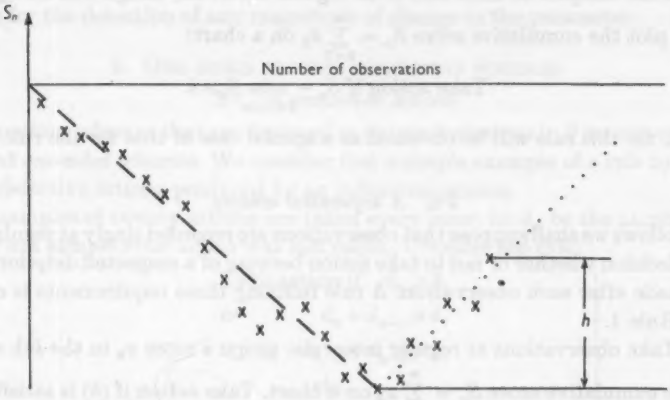


Fig. 1. Control diagram. × × sample points. — — mean path on satisfactory quality. ... mean path on unsatisfactory quality.

An approximation to the distribution of the run length, l, may be found when $P(0) \doteqdot 1$, i.e. when the number of acceptance tests before action is taken is large. If l_1 is the total number of observations in the acceptance tests its characteristic function (C.F.) is

$$\mathscr{E}(e^{tl_1}) = \frac{1-P(0)}{1-P(0)\,\phi_1(t)}, \tag{10}$$

where $\phi_1(t)$ is the C.F. of the number of observations in a single acceptance test. By considering the dominant terms in the repeated differentiation of (10) it follows that, for $P(0)$ near 1, the moments of $l_1/N_1(0)$, and hence those of $l/N_1(0)$, are approximately those of a geometric distribution with mean $P(0)/\{1-P(0)\}$. Hence

$$\Pr(l \leqslant n) \doteqdot 1-\{P(0)\}^{v+1}, \tag{11}$$

where $$v = n/N_1(0).$$

If two independent rules of this type are operated simultaneously (as, for example, when it is desired to control both mean and variance of a normal population for one-sided deviations) the run length of the combination is distributed as the minimum of two independent run lengths, l and l^*; using the approximation (11) the A.R.L. of the combined rules will be λ, where

$$\lambda^{-1} = L^{-1} + L^{*-1}. \tag{12}$$

This relation can be easily seen by considering the stops due to the two rules as occurring at random at rates L and L^*.

In order to calculate the A.R.L. for the transition scheme, the score assigned to each sample may be regarded as a single observation; the average number of samples is therefore given in terms of the characteristics of the appropriate Wald test by equation (9).

2·3. *An integral equation for the* A.R.L.

In general the operating characteristic (O.C.) and average sample number functions of a Wald test are the solutions of integral equations of Fredholm's type. For example, the integral equation for the average sample number $N(Z)$ of a Wald test with boundaries (a, b) and initial score Z is

$$N(Z) = 1 + \int_a^b N(y)\, f(y - Z)\, dy.$$

Using (9) it is therefore possible to calculate the A.R.L. of a general scheme as the ratio of the solutions of two integral equations. We can, however, derive a single integral equation satisfied by the A.R.L., and to do this we introduce an extension to Rule 2.

Rule 3. Take observations and assign scores as for Rule 2. Take action if either

$$(a) \quad S_n \geqslant h \quad \text{and} \quad S_i > 0 \quad \text{for } i = 1, 2, \ldots, n-1,$$

or

$$(b) \quad S_n - \min_{0 \leqslant i < n} S_i \geqslant h,$$

where

$$S_0 = Z \quad (0 \leqslant Z < h) \quad \text{and} \quad S_n = S_{n-1} + x_n. \tag{13}$$

The rule modifies Rule 2 only near its start. A formulation similar to (6) shows that a Wald test with boundaries $(0, h)$ and initial score Z is first applied; if it ends in acceptance, Rule 1 then operates.

Let the distribution function of a single score, x, be $F(x)$, and let $L(Z)$ be the A.R.L. of Rule 3. The A.R.L. of Rule 2 is clearly $L(0)$. By considering the consequences of the first observation and taking expectations we obtain the equation, holding for $0 \leqslant Z < h$,

$$L(Z) = 1 + L(0)\, F(-Z) + \int_0^h L(x)\, dF(x - Z). \tag{14}$$

On substituting for $L(0)$, this equation reduces to the Fredholm form, although the convolution kernel is lost.

2·4. *General remarks*

In the last two sections we have been studying the expected values of a random variable, $l(S, m)$, the number of observations required to satisfy (5) when S is the cumulative score and m the previous minimum. It is plain that $l(S, m)$ is a Markov process in two dimensions where, in general, the states to which the process may move at the next observation are (S_1, m) and (m_1, m_1), where $S_1 \geqslant m$ and where $m_1 < m$. When part of the information is suppressed, e.g. the value of m, the resulting one-dimensional process is no longer Markovian, but this property is regained if the scoring system is that of (6). The states are now defined by the S' score. It follows that every state S' is a regeneration point (Kendall, 1951; Lindley, 1951) of the process. It may also be regarded as a random walk between an absorbing and a 'holding' barrier; when $S'_{n-1} + x_n < 0$, the particle is held at zero until the next observation is taken. The A.R.L. is the mean absorption time, i.e. the mean first passage time to the states $S' \geqslant h$.

3. Two-sided schemes

3·1. *Simultaneous application of two one-sided schemes*

The schemes of § 2 are adequate for the detection of changes in the quality number in one direction only. It is often necessary, e.g. in controlling the mean of a normal population, to detect changes in either direction. A procedure immediately suggesting itself from the previous work is to use two one-sided schemes, one to detect a decrease, and the other an increase, in the quality number. The rule is then:

Rule 4. Take action after the nth article sampled if either

$$S_n - \min_{0 \leqslant i < n} S_i \geqslant h \quad \text{or} \quad \max_{0 \leqslant i < n} S_i - S_n \geqslant k. \tag{15}$$

When this rule is specified by the two co-ordinates $r_n = S_n - \min S_i$, $s_n = \max S_i - S_n$ the run length $l(r, s)$ is a two-dimensional Markov process and the A.R.L. is the mean first passage time to the states (r, s) when $r \geqslant h$ or $s \geqslant k$. An integral equation for the A.R.L. may be derived by introducing a scheme similar to Rule 3, but the equation is more complicated than (14). Alternatively, simultaneous integral equations for $L(r, 0)$ and $L(0, s)$ may be written down in terms of the characteristics of certain Wald tests and the distributions of the overlap of the boundaries at termination. In view of these complications we consider an alternative scheme.

3·2. *A two-sided Wald test*

Barnard (1947) has suggested a sequential test for discriminating between more than two hypotheses, and special cases have been considered by Armitage (1947, 1950) and Sobel & Wald (1949). In this section we specify the particular case needed for the process inspection scheme of § 3·3.

Consider the simultaneous application of two Wald tests, one with initial score on the acceptance boundary and the other with initial score on the rejection boundary. This procedure may be represented graphically as in Fig. 2. The test continues until both the simple Wald tests are ended.

Let the probability that the test ends in the region H_i be $P(H_i)$. Then Sobel & Wald have shown that $P(H_1)$ = O.C. for the first Wald test $= P'$, say. Similarly $P(H_3) = 1 - P''$, so that $P(H_2) = P'' - P'$. In the general case the average sample number (A.S.N.) of the two-sided test is difficult to obtain, but when two boundaries pass through the origin it may be expressed simply in terms of the A.S.N.'s of the composing Wald tests.

Let n, n' and n'' be the number of observations required for a decision in the two-sided test and the simple Wald tests respectively in a realization and let their expectations be denoted by the capital letter with the corresponding dashes.

Then
$$\begin{aligned} n &= \max(n', n'') \\ &= \max(n', n'') + \min(n', n'') - 1, \end{aligned}$$

since the first observation causes at least one of the simple tests to terminate.

Hence
$$n = n' + n'' - 1,$$

and therefore
$$N = N' + N'' - 1. \tag{16}$$

This equation may also be derived from the integral equations by considering the result of the first observation. The simplification leading to (16) arises because the composing tests are independent after the first observation.

3·3. *A scheme based on the two-sided test*

A two-sided process inspection scheme can now be simply specified in terms of the test T, say, described in § 3·2.

Rule 5. Apply T: if it ends in region H_2, reapply T; otherwise take action.

As for (9) we obtain the A.R.L. $$L = N/\{1-P''+P'\}. \tag{17}$$

This scheme is a modification of that based on Rule 4 in which suspected fluctuations in quality are investigated in one direction at a time. In the special case where all the boundaries are horizontal the rule can be stated in terms of maximum and minimum cumulative scores as for Rule 1, making application especially easy; even in the more general case acceptance and rejection scores can be recorded beforehand (cf. Wald, 1947, p. 93), so that only little calculation is necessary when carrying out the rule. The probability that the scheme ends in the H_2 region is $$(1-P'')/(1-P''+P'), \tag{18}$$ so that curves showing the probability that the scheme will indicate the correct direction of the change may be computed from the O.C.'s of the composing Wald tests.

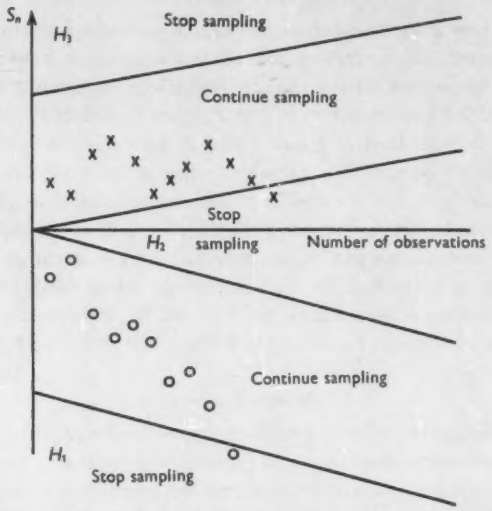


Fig. 2. A two-sided sequential test. ○ ○ sample path ending in H_1. × × sample path ending in H_2.

4. Miscellaneous remarks on process inspection

4·1. *Modifications of the rules*

When one of the above rules has indicated a change in the quality of the output it may be desired to confirm the suspicions by an increased rate of testing. A possible extension to Rule 2 is as follows:

Inspect a fraction f of the output and apply Rule 2 until condition (5) is satisfied; then inspect a fraction f^* until the cumulative score S_n rises a further h_1 or falls h_2. In the former case, take rectifying action, and in the latter, resume inspecting a fraction f and reapply Rule 2.

Another requirement may be to inspect the run of output that leads to the action being taken; this would entail holding production in 'bond' until it became clear that it was satisfactory. For example, with Rule 2, the output would be in bond until the current S_n was the minimum; the average amount held would be $N_1(0)/f$ while production continued and $N_2(0)/f$ when the action was taken.

In practice it might be inconvenient to examine single articles at frequent intervals so that sequential inspection schemes (in the sense of § 2·1) will not be suitable. In such a case a scheme of the transition type may be applied and the size of sample chosen as small as conveniently possible. An alternative procedure is to inspect a fixed number of articles on each occasion but to apply a sequential rule as each article is examined; a decision to take action will be delayed by an average time equivalent to about half the sample size. Another method of grouping is to inspect articles on each occasion until a specified number of the composing Wald tests have terminated or until action is required, whichever is the sooner. The fraction of output sampled would then be a function of the quality number and, for a given θ, could be adjusted by the choice of scheme.

4·2. *Repeated tests*

It has been shown how Rule 2 is equivalent to the repeated application of Wald tests with initial score on the acceptance boundary, and the A.R.L. has been evaluated using that fact. Process inspection schemes can be obtained by repeatedly applying general Wald tests and their A.R.L.'s found in the same manner. When the quality is uniform the behaviour of such schemes is comparable with that of Rules 2 and 3. There is a difference, however, when a deterioration in quality occurs; it is not clear what is the effect on the A.R.L. should an abrupt change in θ take place in the middle of one of the tests. When the tests have initial score on the acceptance boundary we have that $L(Z) \leqslant L(0)$ for Rules 2 and 3 so that the average number of observations until the action is taken is certainly no greater than the A.R.L. of Rule 2 on the new quality. No such statement can be made for more general tests.

Other process inspection schemes may be obtained by the repeated application of different tests or sequences of tests; for example, the control chart method is that of repeated fixed sample size tests.

4·3. *Estimation problems*

If it is required to estimate θ when a process inspection scheme has indicated a change, in practice it may be most convenient to take a further sample for this purpose. Alternatively, an estimator of sorts can be obtained by equating the number of observations in the rejection test to the appropriate conditional A.S.N. function, $N'(\theta)$, say. If there is no root of this equation the estimate can be θ^*, where $N'(\theta^*)$ is the maximum of $N'(\theta)$; if there are several roots, which will be the case in general, some rule such as 'take the largest root' can be used.

Since a process inspection scheme eventually suggests a change in θ whether one has occurred or not a rule for estimating the position of the change may be misleading. If it is known that a change has occurred somewhere within a set of n observations Lindley's Method of Minimum Unlikelihood (1952) will provide an estimator. If it is further known that the change is from θ_0 to θ_1 a simple unbiased estimator, not necessarily integral or within the range $(0, n)$, may be obtained from the meet of the mean paths from zero and $\sum_1^n x_i$, i.e. we take

$$n = \left\{n\theta_1 - \sum_1^n x_i\right\} \Big/ \{\theta_1 - \theta_0\}. \tag{19}$$

In this case the maximum-likelihood estimator is the minimum (maximum if $\theta_0 > \theta_1$) of the sample path for Rule 1 when the scoring is that of the sequential-likelihood ratio test,

$$z_i = \log \frac{f(x, \theta_1)}{f(x, \theta_0)}. \tag{20}$$

5. Process inspection for fractions defective

5·1. *Average run length*

In the case where the articles produced are classified as defective or non-defective explicit formulae for the O.C. and A.S.N. of the Wald test have been given by Burman (1946) and Walker (1950). Let the score for a non-defective article be $-a$ and that for a defective b, where a and b are integers.† The process inspection scheme is given by Rule 2.

The difference equations for the A.R.L. to which the integral equation (14) reduces in this case may be solved by a series transform method as in Walker's paper. We shall, however, use relation (9) and Burman's formulae in which $a = 1$ and it will be convenient to take the formulae in the limit as the fraction defective, $p \to 0$, $b \to \infty$ and $pb = X$. It is necessary to modify these formulae, which do not hold on the boundary, to obtain $P(0)$, $N(0)$. We have, by taking expectations conditional upon the first observation,

$$P(0) = q + pP(b) \tag{21}$$

and

$$N(0) = 1 + pN(b). \tag{22}$$

In the limit the A.R.L. is given by

$$\frac{L(0)}{b} = \frac{1 + XN(b)/b}{X[1 - P(b)]}. \tag{23}$$

5·2. *Choice of scheme*

If the sampling fraction is constant, the A.R.L. is proportional to the average number of articles produced before the inspection scheme causes action to be taken. In some applications it will be more important that stoppages when the quality is good be very infrequent while a fairly lengthy run of poor quality can be tolerated; in other cases a rapid action for bad quality is of prime importance. Sequential schemes to meet these requirements can be selected; e.g. for the first situation we choose the sequential scheme with specified A.R.L. on θ_0 (good quality) and minimum A.R.L. on θ_1 (bad quality). From (23), $pL(0)$ is a function of $X = pb$; consequently, for each h we can find X, and hence b, such that the A.R.L. is L_0 when the fraction defective is p_0. The A.R.L. function for a few values of h may be calculated using the tables of Wald schemes given by Anscombe (1949); for the selection of schemes to the above requirements more extensive tables are needed; these are shown in the Appendix.‡ $pL(0)$ is tabulated as a function of X for $h/b = 1{\cdot}25\,(0{\cdot}25)\,5{\cdot}00$. For $h/b \leqslant 1$, $pL(0) \equiv 1$. As an example, the scheme with $h/b = 2{\cdot}75$ and A.R.L. 4000 at $p_0 = 0{\cdot}02$ has $p_0 L(0) = 80$ at $X \doteqdot 0{\cdot}35$. Hence the required scoring is $b = 17{\cdot}5$. The A.R.L. at $p = 0{\cdot}05$ is given by $p_1 L(0) = 11{\cdot}6$, i.e. $L(0) = 232$ from the value $X_1 = 0{\cdot}875$. In order to select schemes

† With this scoring a rise in the mean sample path corresponds to a rise in the fraction defective. Those who prefer to think in terms of a fall in the standard of production will wish to reverse the signs of the scores.

‡ I am indebted to the Director of the Mathematical Laboratory, University of Cambridge, for permission to use EDSAC for the preparation of these tables.

with given A.R.L. at p_0 and maximum or minimum A.R.L. at p_1 the schemes for different h/b can be compared and the appropriate value chosen.

It has been remarked (see § 2·2) that the process inspection scheme may be carried out as tests (Wald, 1947); it will, however, probably be found more convenient to record the cumulative score S_n on a chart and take action whenever $S_n - \min S_i \geqslant h$. As only the difference between S_n and the minimum is of importance only this information need be entered on a new chart when the old one is completed. If the old chart is withdrawn when the current cumulative score is the minimum the new chart may be started afresh.

5·3. *Comparison of sequential and simple sampling schemes*

The single sampling scheme in which a fixed number, N, of articles is inspected on each occasion and action is taken if c or more defectives are observed has A.R.L. L^*, where

$$L^* = N \Big/ \sum_{i=c}^{N} \binom{N}{i} p^i q^{N-i}. \tag{24}$$

For p small and N large the Poisson approximation may be used; Molina's tables (1947) will be found useful for selecting the best scheme of this type (Page, 1954). In Table 1 the A.R.L. function is given for two values of the argument, p, for five single sampling schemes and in Table 2 the corresponding quantities for five sequential schemes taken from the Appendix table; the first three schemes have specified A.R.L. at $p = 0{\cdot}01$ and minimum A.R.L. at $p = 0{\cdot}03$ and the last two specified A.R.L. at $p = 0{\cdot}03$ and maximum A.R.L. at $p = 0{\cdot}01$.

Table 1. *Single sampling schemes*

N	c	$L^*(0{\cdot}01)$	$L^*(0{\cdot}03)$
63	3	2,500	214
103	4	5,000	275
150	5	8,000	321
70	3	2,050	200
165	5	6,300	300

Table 2. *Sequential schemes*

h/b	b	$L(0{\cdot}01)$	$L(0{\cdot}03)$
2·75	60	2,500	180
3·75	65	5,000	220
4·25	70	8,000	263
3·00	55	3,600	200
3·75	50	13,000	300

For the first three schemes $L(p = 0{\cdot}03) \simeq 0{\cdot}8L^*(p = 0{\cdot}03)$, while the difference in the A.R.L.'s at $p = 0{\cdot}01$ for the last two is proportionately rather more.

It will, however, be noticed that if a deterioration in quality much worse than $p = 0{\cdot}03$ occurs the A.R.L. of the single sampling schemes is at least equal to the sample size while the lower limit of the A.R.L. of the sequential schemes is the largest integer less than $R + 1$. Accordingly the sequential schemes have the advantages of single sampling schemes using small samples for the detection of large changes in quality while retaining a desirable A.R.L. for small changes. These two requirements are in conflict for the choice of the best sample size for single sampling schemes; the best sample size becomes large as the magnitude of the change decreases.

The binomial process inspection scheme may be applied to continuous experiments designed to increase the rate of occurrence of a desirable but rare property. An analogous problem has been considered by Fisher (1952).

6. Applications to storage problems

6·1. *General remarks*

Storage and queuing problems have received attention in a number of recent papers (e.g. Kendall, 1951; Lindley, 1952; Smith, 1953). Many of the problems considered in these papers have placed no restriction on the capacity or content of the store except that the system should be in the equilibrium state where the distribution of input and output are such that the content of the store does not increase without limit. In this section some problems concerned with stores of finite capacity will be discussed briefly.

6·2. *A storage problem*

Consider a store of total capacity h; at certain times ('movement epochs') goods of amount x are sent out and, at the same time, an amount y is received into the store. If the previous contents of the store are insufficient to meet the demand x, the deficiency is made up as far as possible from the new input. If the store would be left overfilled after a movement epoch some of the input is rejected so that the store is left just filled. Let S_n be the store content after the nth movement epoch.

Then
$$\left.\begin{aligned} S_n &= \min(h, S_{n-1}+y-x), \\ S_0 &= Z, \end{aligned}\right\} \tag{25}$$

where Z is the initial content of the store.

We are interested in the occasions when the store is empty and the demand cannot be satisfied; that is to say, when S_n becomes negative. The situation is that of the process inspection Rule 3 to detect a decrease in the quality number when the initial score is Z. The A.R.L. is the average number of movement epochs before the store becomes empty.

We have
$$L(Z) = N(Z) + \{1 - P(Z)\} L(h). \tag{26}$$

The probability that the store becomes empty before becoming full is $P(Z)$, and the expected number of movement epochs between successive epochs of shortage (i.e. when the store is empty) is $N_1(0)$. The number of consecutive epochs of shortage is a geometric variable with parameter $\varpi = \int_{-\infty}^{0} f(z)\,dz$, where $f(z)$ is the frequency function of $z = y - x$.

If any unsatisfied demand is allowed to be supplied out of succeeding inputs, so that, for $S_n < 0$, $|S_n|$ is the total demand outstanding, the average number of movement epochs until the store becomes full is the A.S.N. of the Wald test with a single boundary at h and initial score Z.

6·3. *Stores with a steady drain*

Many problems of practical interest involve stores from which there is a steady output, e.g. storage of fuel for blast furnaces (this has been the subject of some at present unpublished work by H. Herne and D. G. Nikolls). If the store is allowed to accept input whenever the content of the store is $< h$ (even if the input causes the new total content to exceed h) but must reject otherwise, the expected time to an overflow can be written down as before. To find the expected time to the exhaustion of the store we need, in general, to create a regeneration point after the store has overflowed; for example, we can insert the condition that the next input after an overflow shall arrive when the content is Z_0. The previous methods may then be applied.

When the upper limit, h, of the store may never be exceeded the problem is more complicated. The net increment between inputs is no longer sufficient to describe the behaviour of the store content. In the Wald test it is necessary to take two separate steps, first the amount of input and then the drain. The simple theory can be applied in some special cases. If the amount of input on each occasion is fixed (and equal to unity, say) and the drain between input epochs is a random variable, y, a Wald test with boundaries $(0, h-1)$ and increments $1-y$ may be used to construct the theory as before. Similarly, a Wald test with boundaries $(1, h)$ and increments $x-1$ may be used when there are inputs of amount x at unit time intervals. When the random variable has the exponential distribution an explicit solution for the characteristics of the Wald test is available (Anscombe & Page, 1954).

7. Deferred sentencing schemes

7·1. *General remarks*

Process inspection schemes are designed to detect variations in the quality of output as it is being produced so that faults in the process may be traced promptly. In some cases it is impracticable to apply such a scheme, but after production it is necessary that runs of good and bad quality be separated, i.e. the output must be sentenced. Different action will then be taken with the two qualities. In the original formulation the problem is that of identifying the subsamples with the different parameter values. Some deferred sentencing schemes for articles classed as defective or not have been given by Anscombe *et al.* (1947); in these schemes articles are sampled and tested one by one and rules for the rejection of output are based on the occurrence of d defectives in any sequence of N articles examined. Roughly speaking, these moving average rules accept the output until a deterioration in quality is noticed and then reject until an improvement appears. Similar schemes can be constructed from process inspection schemes.

7·2. *Repeated process inspection schemes*

When a deterioration in quality is shown by a deviation in one direction (an increase, say) of the quality number, a deferred sentencing scheme may be obtained by applying two process inspection schemes consecutively. First, apply Rule 2 to detect an increase in θ; let l be the position of the minimum S_i when the rule operates. Then apply Rule 2 to detect a decrease in θ; let m be the position of the maximum S_i. Reject articles $l+1$ and m and all output within those limits. Because there is probability one that the process inspection scheme operates even if the quality, θ, remains constant some of the output will be rejected by this deferred sentencing scheme whatever the value of θ, satisfactory or not. In the notation of § 2·2 with $P^*(Z)$, etc., written for the characteristics of the Wald test with boundaries $(-k, 0)$, an estimate of the proportion of output rejected is

$$\frac{N_2(0)+\frac{\{1-P^*(0)\}}{P^*(0)}N_2^*(0)}{L(0)+L^*(0)}. \tag{27}$$

Two-sided deferred sentencing schemes may be obtained in a similar manner. For example, a scheme could be:

Apply Rule 5 until it operates; suppose that it does so because of a hit on the $H_1(H_3)$ boundary (thus suggesting a decrease (increase) in θ). Then apply Rule 1 to detect an increase (decrease) in θ; when it operates repeat the cycle.

The schemes suggested above may be regarded as a succession of Wald tests with the rule that all output leading to a rejection test is rejected. Other schemes may be constructed using different sequences of tests.

8. RECTIFYING SCHEMES

8·1. *General remarks*

The main purpose of rectifying inspection is to check that the output maintains a given standard and to improve the quality by replacing defective articles if necessary. Often it is require to ensure that the average quality of the product after inspection (the average outgoing quality: A.O.Q.) is better than some specified standard whatever be the quality before inspection. We suppose that articles are either defective or non-defective. A fraction (possibly large) of the output is inspected and any defective articles are rectified or replaced by non-defectives. Several schemes for the rectifying inspection of continuous output have been suggested, notably those of Dodge (1943) and Wald & Wolfowitz (1945). In these schemes a fraction, $f < 1$, of the output is inspected until there are signs that the quality is not as good as that required; 100 % inspection is then started and continued until an improvement is noticed, when the cycle is repeated. We consider schemes in which the signs of improvement or deterioration are based on the results of tests repeatedly applied.

8·2. *Repeated tests*

Suppose that we are given two statistical tests, T and T^*, for the quality of the production and that each has two possible results, acceptance and rejection. Then we consider the rule:

'Inspect a fraction f of the output and apply test T: if it accepts, reapply T; if it rejects, start 100 % inspection and apply T^*. If T^* rejects, continue 100 % inspection and reapply T^*; if it accepts, inspect a fraction f and restart the cycle of tests.'

When the production process is under control so that the proportion of defectives, p, produced remains constant we can express the A.O.Q. in terms of the A.R.L.'s L and L^*, of the process inspection schemes obtained by repeatedly applying T and T^* respectively. We have

$$\text{A.O.Q.} = \frac{pL(1-f)}{L+fL^*}. \tag{28}$$

The fraction of production inspected, $I(p)$, is given by

$$I(p) = \frac{f(L+L^*)}{L+fL^*}. \tag{29}$$

More complex rules can be formulated in which more than two tests are used or in which succeeding tests depend upon characteristics of preceding ones. Tests allowing more than two decisions may be used to cater for multi-dimensional inspection.

8·3. *Special cases*

The scheme suggested by Dodge is:

'Inspect a fraction f of the output until a defective occurs; then do 100 % inspection until a run of i non-defectives is observed.' In this case the test T may be taken as that based on a fixed sample size of one article with acceptance for a non-defective and rejection for a defective. T^* is a binomial sequential test with boundaries $(0, i)$, starting score i and penalty for a defective $\geqslant i$.

This scheme bases its decision to start 100 % inspection on the appearance of one defective only; if this treatment of a first offender seems harsh it may be preferred to use for T a sequential test with initial score zero and boundaries $(0, h)$, where h is greater than the penalty for a defective.

Schemes may be obtained using Wald tests with general starting scores, but disadvantages similar to those given for process inspection schemes are present when the quality is not constant; these same disadvantages will apply in greater or less degree to the use of all tests with an A.S.N. much greater than one.

For the sampling plan A of Wald and Wolfowitz T is a sequential test with only a rejection boundary and T^* a fixed size test with only an acceptance criterion. The position of the boundary in the next test T and the number of observations in T^* depend on the overlap on the boundary in the preceding test.

I wish to express my thanks to Mr F. J. Anscombe, who suggested this subject of research and supervised it, and to Dr D. R. Cox, Mr D. V. Lindley and Dr J. Wishart for reading the draft of the paper. I also wish to thank the Department of Scientific and Industrial Research for the award of a maintenance grant.

REFERENCES

Anscombe, F. J. (1949). *J. R. Statist. Soc.* A, **62**, 180.
Anscombe, F. J., Godwin, H. J. & Plackett, R. L. (1947). *J. R. Statist. Soc. Suppl.* **9**, 198.
Anscombe, F. J. & Page, E. S. (1954). *Biometrika*, **41**, 252.
Armitage, P. (1947). *J. R. Statist. Soc. Suppl.* **9**, 250.
Armitage, P. (1950). *J. R. Statist. Soc.* B, **12**, 145.
Aroian, L. A. & Levene, H. (1950). *J. Amer. Statist. Ass.* **45**, 520.
Barnard, G. A. (1947). *J. Amer. Statist. Ass.* **42**, 658.
Bartlett, M. S. (1953). *Proc. Camb. Phil. Soc.* **49**, 263.
Burman, J. P. (1946). *J. R. Statist. Soc. Suppl.* **8**, 98.
Dodge, H. F. (1943). *Ann. Math. Statist.* **14**, 264.
Dudding, B. P. & Jennet, W. J. (1942). British Standards Institution, B.S. 600R.
Duncan, A. J. (1952). *Quality Control and Industrial Statistics*. Chicago: Irwin.
Fisher, R. A. (1952). *Biometrics*, **8**, 183.
Herne, H. & Nickolls, G. D. Unpublished.
Kendall, D. G. (1951). *J. R. Statist. Soc.* B, **13**, 184.
Lindley, D. V. (1951). *J. R. Statist. Soc.* B, **13**, 173.
Lindley, D. V. (1952). *Proc. Camb. Phil. Soc.* **48**, 277.
Lindley, D. V. (1953). *J. R. Statist. Soc.* B, **15**, 30.
Molina, E. C. (1947). *Poisson's Exponential Binomial Limit*. New York: Von Nostrand.
Neyman, J. & Pearson, E. S. (1936). *Statist. Res. Mem.* **1**, 113.
Page, E. S. (1954). *J. R. Statist. Soc.* B, **16**, (at Press).
Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*. New York: Macmillan.
Smith, W. L. (1953). *Proc. Camb. Phil. Soc.* **49**, 449.
Sobel, M. & Wald, A. (1949). *Ann. Math. Statist.* **20**, 502.
Wald, A. (1947). *Sequential Analysis*. New York: Wiley.
Wald, A. & Wolfowitz, J. (1945). *Ann. Math. Statist.* **16**, 30.
Walker, A. M. (1950). *J. R. Statist. Soc.* B, **12**, 301.

h/b / $X = pb$	1·25	1·50	1·75	2·00	2·25	2·50	2·75	3·00	3·25	3·50	3·75	4·00	4·25	4·50	4·75	5·00
0·05	28·1	41·0	77·4	829	$1{\cdot}41\times10^3$	$2{\cdot}94\times10^3$	$9{\cdot}14\times10^3$	$4{\cdot}64\times10^4$	$1{\cdot}01\times10^5$	$2{\cdot}77\times10^5$	$1{\cdot}03\times10^6$	$3{\cdot}20\times10^6$	$8{\cdot}15\times10^6$	—	—	—
0·10	14·8	21·0	37·7	216	354	688	$1{\cdot}79\times10^3$	$5{\cdot}65\times10^3$	$1{\cdot}14\times10^4$	$2{\cdot}71\times10^4$	$7{\cdot}47\times10^4$	$1{\cdot}81\times10^5$	$4{\cdot}19\times10^5$	$1{\cdot}07\times10^6$	$2{\cdot}71\times10^6$	$6{\cdot}28\times10^6$
0·15	10·4	14·4	31·1	100	158	290	660	$1{\cdot}65\times10^3$	$3{\cdot}02\times10^3$	$6{\cdot}58\times10^3$	$1{\cdot}52\times10^4$	$3{\cdot}20\times10^4$	$6{\cdot}65\times10^4$	$1{\cdot}46\times10^5$	$3{\cdot}19\times10^5$	$6{\cdot}67\times10^5$
0·20	8·13	11·1	18·3	59·1	90·3	156	321	686	$1{\cdot}21\times10^3$	$2{\cdot}30\times10^3$	$4{\cdot}80\times10^3$	$9{\cdot}18\times10^3$	$1{\cdot}75\times10^4$	$3{\cdot}46\times10^4$	$6{\cdot}78\times10^4$	$1{\cdot}30\times10^5$
0·25	6·80	9·13	14·5	39·7	58·9	97·0	183	351	588	$1{\cdot}06\times10^3$	$1{\cdot}96\times10^3$	$3{\cdot}46\times10^3$	$6{\cdot}14\times10^3$	$1{\cdot}11\times10^4$	$2{\cdot}00\times10^4$	$3{\cdot}56\times10^4$
0·30	5·92	7·81	12·0	29·1	41·8	66·0	116	204	326	548	941	$1{\cdot}56\times10^3$	$2{\cdot}60\times10^3$	$4{\cdot}39\times10^3$	$7{\cdot}37\times10^3$	$1{\cdot}23\times10^4$
0·35	5·29	6·88	10·3	22·5	31·6	47·9	79·1	130	199	316	509	799	$1{\cdot}26\times10^3$	$2{\cdot}00\times10^3$	$3{\cdot}16\times10^3$	$4{\cdot}97\times10^3$
0·40	4·82	6·18	9·04	18·2	25·0	36·5	57·3	89·3	138	198	301	450	674	$1{\cdot}02\times10^3$	$1{\cdot}53\times10^3$	$2{\cdot}28\times10^3$
0·45	4·45	5·65	8·07	15·3	20·4	29·0	43·5	64·6	91·2	132	192	274	303	565	810	$1{\cdot}16\times10^3$
0·50	4·16	5·22	7·31	13·1	17·2	23·7	34·2	48·8	66·7	92·9	130	178	245	388	465	638
0·55	3·93	4·87	6·71	11·5	14·8	19·9	27·8	38·3	50·8	68·4	92·0	122	162	216	286	377
0·60	3·73	4·59	6·21	10·2	12·9	17·0	23·1	30·9	40·0	52·3	68·1	87·8	113	145	186	238
0·65	3·56	4·34	5·80	9·21	11·5	14·9	19·7	25·7	32·5	41·3	52·4	65·7	82·2	103	128	158
0·70	3·42	4·14	5·46	8·42	10·4	13·2	17·1	21·8	27·0	33·6	41·6	50·9	62·2	75·7	91·7	111
0·75	3·30	3·96	5·17	7·77	9·47	11·8	15·1	18·8	22·9	28·0	33·9	40·7	48·7	58·0	68·7	81·1
0·80	3·19	3·81	4·91	7·23	8·72	10·8	13·5	16·6	19·9	23·8	28·4	33·5	39·3	45·9	53·4	61·8
0·85	3·10	3·67	4·69	6·78	8·10	9·88	12·2	14·8	17·5	20·7	24·2	28·2	32·6	37·5	42·8	48·8
0·90	3·01	3·55	4·50	6·40	7·57	9·15	11·1	13·3	15·6	18·2	21·1	24·2	27·6	31·4	35·4	39·7
0·95	2·94	3·45	4·33	6·07	7·13	8·53	10·3	12·2	14·1	16·3	18·7	21·2	23·9	26·8	30·0	33·3
1·00	2·87	3·35	4·19	5·78	6·75	8·01	9·56	11·2	12·9	14·7	16·7	18·8	21·1	23·4	25·9	28·5
1·05	2·81	3·27	4·05	5·54	6·42	7·56	8·96	10·4	11·9	13·5	15·2	17·0	18·8	20·8	22·8	24·9
1·10	2·76	3·19	3·93	5·32	6·13	7·18	8·44	9·75	11·0	12·5	14·0	15·5	17·1	18·7	20·4	22·1
1·15	2·71	3·12	3·83	5·13	5·88	6·85	8·00	9·18	10·3	11·6	12·9	14·3	15·6	17·0	18·5	20·0
1·20	2·67	3·06	3·73	4·96	5·66	6·56	7·62	8·70	9·75	10·9	12·1	13·2	14·5	15·7	17·0	18·3
1·25	2·63	3·00	3·64	4·81	5·48	6·30	7·28	8·28	9·24	10·3	11·3	12·4	13·5	14·6	15·7	16·9
1·50	2·47	2·78	3·31	4·26	4·75	5·38	6·12	6·85	7·53	8·26	8·98	9·72	10·4	11·2	11·9	12·7
1·75	2·36	2·62	3·08	3·91	4·31	4·89	5·44	6·03	6·58	7·16	7·75	8·31	8·91	9·48	10·1	10·6
2·00	2·28	2·50	2·92	3·68	4·02	4·46	5·00	5·52	5·99	6·48	7·00	7·50	7·98	8·50	9·00	9·48
2·25	2·22	2·42	2·79	3·52	3·80	4·21	4·70	5·18	5·58	6·03	6·50	6·95	7·38	7·85	8·30	8·75
2·50	2·18	2·35	2·70	3·40	3·65	4·02	4·48	4·93	5·30	5·72	6·15	6·55	6·97	7·40	7·82	8·22
2·75	2·14	2·30	2·63	3·32	3·54	3·83	4·32	4·74	5·09	5·47	5·91	6·27	6·63	7·04	7·45	7·84
3·00	2·12	2·26	2·56	3·25	3·44	3·76	4·20	4·61	4·92	5·31	5·70	6·06	6·42	6·81	7·20	7·56

ON NAHORDNUNG AND FERNORDNUNG IN SAMPLES OF LITERARY TEXTS

By WILHELM FUCKS

Physikalisches Institut, Aachen

1. Introduction

In a previous paper (Fucks, 1952, cited as I), several possibilities were outlined to describe the style of samples of literary texts by means of mathematical characteristics. The computing of practical examples, however, was confined to the most simple ones of these style characteristics. Most of them were derived from the frequency distributions of the properties of elementary units of the text themselves. It was, however, pointed out briefly, that the properties of groups of elementary units of the text can also be used to define characteristic numbers and, as an example, the values of an $\bar{l}_{ik}$-matrix were given.

To explain the aim of this paper, we assume the text to be divided into elements. The elements will be denoted by the numbers $1, 2, \ldots, A$. It is perfectly arbitrary what parts of the text are chosen as elements: syllables, words, parts of the phrases, whole phrases, chapters, etc., or substantives, adjectives, verbs, etc., or cases of substantives, tenses of verbs, etc., or metric elements, and so on. The elements are to be considered as characterized by different marks, one or several of them for each element. These marks (or properties) can be chosen, within certain limits, in an arbitrary manner. However, to clarify our conception with the help of a special example, we shall generally in this paper, as in paper I, choose the words of a text as elements and the numbers of the syllables of the words as the interesting properties of these elements.

Consequently, we shall now assume that the text is broken down into the words; and that these have been mixed together and then picked up, one after the other, perfectly at random and given the numbers $1, 2, \ldots, A$. Theoretically, if we should repeat this procedure often enough, we should obtain the natural order of the original text as one of the different arrangements arising from picking up the words at random. We conclude, therefore, that we do not restrict the generality of our conclusions, if we start by marking the words of the text in their original order by the numbers $1, 2, \ldots, A$.

Now, let us assume that we have computed the values of p_i, the proportion of the A words of a given text having i syllables and also the various characteristic numbers calculated in I, namely, $\bar{\imath}$, the mean number of syllables per word; $S = -k \sum_i p_i \log p_i$, the 'entropy'; and the 'trace', s. Considering these results, we ask how much of the structure of the text is described in an exact manner by these characteristic numbers. The answer is that our characteristic numbers do not tell us much more of the text than the vapour of a mixture of several fluids tells us about the fluid-structures proper, i.e. the real structure of the text does not enter into the characteristic numbers so far studied in detail. What we have taken into account with the characteristic numbers actually computed in I is the choice of a certain number, z, of words out of the reservoir V of the words of the language in which the text is written. The words enter into our characteristics as formed grammatically. But the grammatical forming process is contained in our characteristic numbers only in a rudimentary way, i.e. only in so far as the numbers of the syllables of the words are affected.

It may be regarded as rather noteworthy, that so much can be achieved in characterizing the style of an author, starting our calculus with such scanty material. We have, indeed, until now not used the various moments of the frequency distributions computed in I. In the meantime the moments of the orders two to four and other higher characteristic numbers of the distributions discussed in I have been calculated and the results will be published in another paper.

All the characteristic numbers calculated from properties of the fundamental text elements themselves will hence be called characteristic numbers of the first order. If we calculate, however, a characteristic number using simultaneously the properties of groups of $2, 3, \ldots, n$ fundamental elements at $2, 3, \ldots, n$ different positions in the text, we will describe those characteristic numbers as of the second, third, ..., nth order.

Besides these distinctions another distinction is important. If the groups of elements considered for the computing of a characteristic number of an order higher than one are arranged in the original text in the immediate neighbourhood of one another, we speak of Nahordnung; if not, we speak of Fernordnung. It will be shown that it is advisable to define different types of characteristic numbers for both these cases. Certain characteristic numbers useful for the study of groups of consecutive elements (Nahordnung) prove either mathematically or practically uninteresting when applied to the study of the relations between the properties of distant parts of the text (Fernordnung) and vice versa.

The distinction between Nahordnung and Fernordnung is perfectly unique when applied to characteristic numbers of the second order; applied to characteristic numbers of an order higher than two it is not. There are, obviously, characteristic numbers of an order higher than two that are pure Nahordnungs-cases, and, on the other hand, characteristic numbers of an order higher than two that are pure Fernordnungs-cases, but there exist mixed cases as well.

The aim of the present paper is to define characteristic numbers describing Nahordnungs- and Fernordnungs-relationships of any order. The calculation of examples of Nahordnungs- and Fernordnungs-characteristic numbers, however, will be confined to the second order exclusively.

2. NAHORDNUNG

A sample of a text may be given as follows:

$$\binom{e_1}{f_1}\binom{e_2}{f_2}\binom{e_3}{f_3}\cdots\binom{e_{i-1}}{f_{i-1}}\binom{e_i}{f_i}\binom{e_{i+1}}{f_{i+1}}\cdots\binom{e_N}{f_N},$$

the elements $e_1, e_2, \ldots$ having certain properties $f_1, f_2, \ldots$ respectively. The e_i may be, for instance, assumed to be the words and the f_i to be the numbers of syllables of the words.

We now break up the text into pairs, triplets, ..., n-tuplets of elements as indicated in the following lines (a), (b), (c), etc.:

$$(a) \quad | \; e_1 \; e_2 \; | \; e_3 \; e_4 \; | \; e_5 \; e_6 \; | \; e_7 \; e_8 \; | \; e_9 \; e_{10} \; | \; e_{11} \; e_{12} \; | \; \ldots,$$

$$(b) \quad | \; e_1 \; e_2 \; e_3 \; | \; e_4 \; e_5 \; e_6 \; | \; e_7 \; e_8 \; e_9 \; | \; e_{10} \; e_{11} \; e_{12} \; | \; \ldots,$$

$$(c) \quad | \; e_1 \; e_2 \; e_3 \; e_4 \; | \; e_5 \; e_6 \; e_7 \; e_8 \; | \; e_9 \; e_{10} \; e_{11} \; e_{12} \; | \; \ldots,$$

and consider the properties connected with the first, the second, the third, etc., groups in each line. The arrangement of the groups one after the other may be supposed, for the

moment, to be that of the original text. Thus we find sets of properties bound together by their being found in 2, 3, ... consecutive elements of the text, thus:

$$\begin{array}{lllll} \text{for } n=2: & f_i, & f_{i+1}; & & \\ \text{for } n=3: & f_i, & f_{i+1}, & f_{i+2}; & \\ \quad\dots & \quad\dots & \quad\dots & \quad\dots & \\ \quad\dots & \quad\dots & \quad\dots & \quad\dots & \\ \text{for } n=k: & f_i, & f_{i+1}, & f_{i+2}, \ \dots, & f_{i+k}. \end{array}$$

We now compute the relative frequency distributions of the n-tuplets for the whole of the given text, defining thereby matrices of the order n:

$$p_{ik}, \quad p_{ikl}, \quad \dots, \quad p_{ikl\dots n}.$$

These matrices, and certain numbers that will be derived from them afterwards, prove to be most appropriate characteristic numbers to describe relations between the properties of groups of neighbouring elements.

It will be noticed that the splitting up process of the text into groups as described above is not unique. With binary groups there are two possibilities:

$$| \ e_1 \ \ e_2 \ | \ e_3 \ \ e_4 \ | \ e_5 \ \ e_6 \ | \ e_7 \ \dots \quad \text{and} \quad e_1 \ | \ e_2 \ \ e_3 \ | \ e_4 \ \ e_5 \ | \ e_6 \ \ e_7 \ | \ \dots.$$

Generally, with groups consisting of n elements, there are n different possibilities. If the text is written in prose, there will, in most cases, be no difference with regard to the style characteristics if we break up the text in any of these ways. With a text written in verse, however, there may be a remarkable difference according to the phase of the beginning of the process of subdividing the text, especially if the properties of the elements entering into the characteristic numbers are metric ones.

There are obviously additional possibilities for dividing the text into groups, if we admit combinations of the possibilities described so far. For binary groups, for instance, we can break up the text in both ways described above and calculate the p_{ik}-matrix using the frequency distributions of both sets of groups (hence called distributions from the combined subdivision).

We may ask, to what extent the matrices $p_{ik\dots}$ take account of the very structure of the text. After forming the groups of n elements we may combine these groups and in order to perform the calculations described in the following chapters, we may pick up the groups perfectly at random one after the other and give them consecutive numbers $1, 2, \dots, A$, without changing the matrices $p_{ikl\dots}$ or the characteristic numbers derived from them. The amount to which the structure proper enters our style characteristics is obviously appropriately described by the term Nahordnung of the text.

2·1. *The matrix p_{ik} of the frequency distribution of binary groups*

We assume a text consisting of A elements $e_1, e_2, \dots, e_A$ with the properties $f_1, f_2, \dots, f_A$, respectively. After applying the combined subdivision described above we compute the number A_{ik} of the $A-1$ pairs of consecutive elements, that bear the properties f_i and f_k, respectively. In this way we find the distributions of the relative frequencies of the (i,k)-pairs:

$$p_{ik} = \frac{A_{ik}}{A-1}. \tag{1}$$

The following relations obviously hold for the p_{ik}-matrices:

$$\sum_i \sum_k p_{ik} = 1, \tag{2}$$

$$\sum_k p_{ik} = p_i, \quad \sum_i p_{ik} = p_k. \tag{3}$$

With the words as elements and the number of the syllables per word as the properties, we find for some of the authors and works studied in I the following p_{ik}-matrices:

Table 1. *p_{ik}-matrix for Shakespeare*, Othello

	1	2	3	4	5
1	0·6137	0·1249	0·0392	0·0103	0·0007
2	0·1246	0·0190	0·0074	0·0012	0·0002
3	0·0386	0·0057	0·0025	0·0006	0·0001
4	0·0102	0·0015	0·0004	0·0004	0·0000
5	0·0010	0·0000	0·0000	0·0000	0·0000

Table 2. *p_{ik}-matrix for Rilke*, Cornet

	1	2	3	4	5
1	0·3936	0·1796	0·0396	0·0101	0·0004
2	0·1808	0·1003	0·0186	0·0032	0·0000
3	0·0376	0·0202	0·0020	0·0000	0·0000
4	0·0089	0·0045	0·0000	0·0000	0·0000
5	0·0004	0·0000	0·0000	0·0000	0·0000

Table 3. *p_{ik}-matrix for Hesse*, Steppenwolf

	1	2	3	4	5	6	7
1	0·26556	0·16813	0·05732	0·02019	0·00550	0·00140	0·00023
2	0·16331	0·09495	0·03596	0·01442	0·00324	0·00086	0·00009
3	0·06147	0·03272	0·01190	0·00446	0·00140	0·00023	0·00005
4	0·02068	0·01374	0·00469	0·00306	0·00054	0·00014	0·00000
5	0·00518	0·00324	0·00158	0·00068	0·00014	0·00000	0·00000
6	0·00122	0·00104	0·00023	0·00009	0·00000	0·00000	0·00000
7	0·00018	0·00009	0·00000	0·00005	0·00009	0·00000	0·00000

Table 4. *p_{ik}-matrix for Mann*, Buddenbrooks

	1	2	3	4	5	6	7	8
1	0·2616	0·1620	0·0586	0·0238	0·0055	0·0012	0·0002	0·0001
2	0·1664	0·0982	0·0318	0·0151	0·0034	0·0008	0·0002	0·0001
3	0·0543	0·0355	0·0119	0·0056	0·0013	0·0002	0·0000	0·0001
4	0·0239	0·0157	0·0042	0·0026	0·0010	0·0002	0·0001	0·0000
5	0·0056	0·0037	0·0011	0·0007	0·0003	0·0001	0·0000	0·0000
6	0·0013	0·0008	0·0001	0·0001	0·0001	0·0000	0·0000	0·0000
7	0·0002	0·0001	0·0001	0·0001	0·0000	0·0000	0·0000	0·0000
8	0·0002	0·0001	0·0000	0·0000	0·0000	0·0000	0·0000	0·0000

Table 5. p_{ik}-*matrix for Mann*, Zauberberg

	1	2	3	4	5	6	7	8
1	0·2643	0·1618	0·0590	0·0256	0·0077	0·0020	0·0007	0·0001
2	0·1608	0·0792	0·0348	0·0142	0·0041	0·0010	0·0002	0·0001
3	0·0607	0·0355	0·0127	0·0059	0·0012	0·0003	0·0000	0·0001
4	0·0256	0·0135	0·0065	0·0032	0·0007	0·0001	0·0001	0·0000
5	0·0066	0·0042	0·0023	0·0006	0·0002	0·0000	0·0000	0·0000
6	0·0019	0·0007	0·0006	0·0002	0·0000	0·0000	0·0000	0·0000
7	0·0003	0·0003	0·0002	0·0002	0·0000	0·0000	0·0000	0·0000
8	0·0001	0·0001	0·0000	0·0001	0·0000	0·0000	0·0000	0·0000

Table 6. p_{ik}-*matrix for Carossa*, Geheimnisse des reifen Lebens

	1	2	3	4	5	6	7
1	0·24200	0·16840	0·06492	0·02017	0·00557	0·00040	0·00018
2	0·16857	0·09502	0·03673	0·01509	0·00374	0·00076	0·00022
3	0·06211	0·03745	0·01305	0·00588	0·00169	0·00027	0·00004
4	0·02213	0·01492	0·00387	0·00187	0·00058	0·00004	0·00000
5	0·00646	0·00365	0·00134	0·00045	0·00013	0·00000	0·00000
6	0·00094	0·00045	0·00022	0·00009	0·00004	0·00000	0·00000
7	0·00018	0·00018	0·00004	0·00000	0·00004	0·00000	0·00000

Table 7. p_{ik}-*matrix for Jaspers*, Der philosophische Glaube

	1	2	3	4	5	6	7	8	9
1	0·23877	0·13408	0·06872	0·04215	0·01824	0·00463	0·00078	0·00014	0·00009
2	0·13868	0·06090	0·02758	0·01609	0·00689	0·00141	0·00017	0·00023	0·00000
3	0·06950	0·02822	0·01273	0·00793	0·00282	0·00072	0·00011	0·00003	0·00003
4	0·03962	0·01626	0·00954	0·00523	0·00267	0·00049	0·00016	0·00006	0·00003
5	0·01494	0·00888	0·00422	0·00259	0·00149	0·00049	0·00003	0·00003	0·00000
6	0·00385	0·00198	0·00124	0·00046	0·00029	0·00014	0·00003	0·00003	0·00000
7	0·00060	0·00032	0·00026	0·00003	0·00000	0·00000	0·00000	0·00000	0·00000
8	0·00029	0·00003	0·00003	0·00000	0·00000	0·00000	0·00000	0·00000	0·00000
9	0·00009	0·00000	0·00003	0·00000	0·00003	0·00000	0·00000	0·00000	0·00000

A closer study of the matrices shows remarkable differences with regard to the different works from which they are derived. Certain characteristic factors, which bring out more clearly the differences of the style of the different works, will be defined and computed in the following sections.

2·2. *The moments of the first order of the p_{ik}-matrices*

We study the moments M_1' and M_1'':

$$M_1' = \sum_i \sum_k i p_{ik}, \quad M_1'' = \sum_i \sum_k k p_{ik}. \tag{4}$$

If we consider the p_{ik}-distribution as a distribution of masses in a plane co-ordinate system with the co-ordinates i and k, the values of M_1' and M_1'' give the co-ordinates i^* and k^* of the

mass centre of this mass distribution. With the words as elements and the number of syllables per word as properties, there holds the relation $\sum_i i p_{ik} = \sum_i i p_{ki}$. Hence we obtain

$$M_1' = M_1'' = \bar{\imath}.$$

From the matrices given in Tables 1–7 we compute Table 8.

Table 8. *Mean values of the p_{ik}-matrix*

Author	Work	$M_1' = M_1''$
Shakespeare	*Othello*	1·2943
Rilke	*Cornet*	1·4506
Hesse	*Steppenwolf*	1·7245
Mann	*Buddenbrooks*	1·7376
Mann	*Zauberberg*	1·7567
Carossa	*Geheimnisse*	1·7494
Jaspers	*Der phil. Glaube*	1·8871

There are, however, cases in which the relation mentioned above does not hold, e.g. p_{ik} = relative frequencies of words per sentence and syllables per sentence. Characteristics derived in this way will be described in another publication.

2·3. *The variance*

The study of the higher moments of the p_{ik}-matrices will be confined in this paper to the moments of the second order, i.e. the variance M_2. This moment is given by a symmetrical matrix with three independent matrix elements. If we consider the p_{ik}-matrix as described before as a mass distribution in a plane (i, k)-co-ordinate system, the three elements of the M_2-matrix are the two moments of inertia σ_{11} and σ_{22} with respect to the two axes $i = i^*$ and $k = k^*$, respectively, and the product of inertia $\sigma_{12} = \sigma_{21}$:

$$M_2 = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}, \tag{5}$$

with

$$\left.\begin{aligned} \sigma_{11} &= \sum_i \sum_k (i - M_1')^2 p_{ik}, \\ \sigma_{22} &= \sum_i \sum_k (k - M_1'')^2 p_{ik}, \\ \sigma_{12} = \sigma_{21} &= \sum_i \sum_k (i - M_1')(k - M_1'') p_{ik}. \end{aligned}\right\} \tag{6}$$

In the special case with $\sum_i i p_{ik} = \sum_i i p_{ki}$ we obtain

$$\sigma_{11} = \sigma_{22} = \sum_i i^2 p_i - \bar{\imath}^2.$$

The examples of texts, studied in this paper, lead to the values for the matrix M_2 of Table 9. In computing these values there arise small differences between σ_{11} and σ_{22} due to errors connected with the establishing of the matrices. In Table 9 the mean value of the calculated σ_{11} and σ_{22} values is given.

Table 9. *Variance of the p_{ik}-distributions*

Shakespeare, *Othello*

$$M_2 = \begin{pmatrix} 0{\cdot}3911 & -0{\cdot}0117 \\ -0{\cdot}0117 & 0{\cdot}3911 \end{pmatrix}$$

Rilke, *Cornet*

$$M_2 = \begin{pmatrix} 0{\cdot}4543 & -0{\cdot}0076 \\ -0{\cdot}0076 & 0{\cdot}4543 \end{pmatrix}$$

Hesse, *Steppenwolf*

$$M_2 = \begin{pmatrix} 0{\cdot}8736 & +0{\cdot}0110 \\ +0{\cdot}0110 & 0{\cdot}8736 \end{pmatrix}$$

Mann, *Buddenbrooks*

$$M_2 = \begin{pmatrix} 0{\cdot}9104 & +0{\cdot}0094 \\ +0{\cdot}0094 & 0{\cdot}9104 \end{pmatrix}$$

Mann, *Zauberberg*

$$M_2 = \begin{pmatrix} 0{\cdot}9881 & +0{\cdot}0017 \\ +0{\cdot}0017 & 0{\cdot}9881 \end{pmatrix}$$

Carossa, *Geheimnisse des reifen Lebens*

$$M_2 = \begin{pmatrix} 0{\cdot}8774 & -0{\cdot}0133 \\ -0{\cdot}0133 & 0{\cdot}8774 \end{pmatrix}$$

Jaspers, *Der philosophische Glaube*

$$M_2 = \begin{pmatrix} 1{\cdot}4016 & -0{\cdot}0912 \\ -0{\cdot}0912 & 1{\cdot}4016 \end{pmatrix}$$

We transform the M_2-matrix on principal axes and we find the principal variances σ_{I} and σ_{II}. In our mass distribution they are to be compared with the principal moments of inertia. We obtain their values as the solutions of the quadratic equation in λ:

$$\begin{vmatrix} \sigma_{11}-\lambda & \sigma_{12} \\ \sigma_{21} & \sigma_{22}-\lambda \end{vmatrix} = 0. \tag{7}$$

The results for our examples of text are given in Table 10.

Table 10. *Principal variances of the p_{ik}-distributions*

Author	Work	σ_{I}	σ_{II}
Shakespeare	*Othello*	0·3724	0·4028
Rilke	*Cornet*	0·4467	0·4619
Hesse	*Steppenwolf*	0·8846	0·8626
Mann	*Buddenbrooks*	0·9109	0·9099
Mann	*Zauberberg*	0·9864	0·9898
Carossa	*Geheimnisse*	0·8907	0·8641
Jaspers	*Der phil. Glaube*	1·3104	1·4928

In Fig. 1 we plot σ_{II} against σ_{I}.

2·4. *The correlation in the region of Nahordnung*

We compute, using the M_2-matrix, the correlation coefficient c:

$$c = \frac{\sigma_{12}}{\sqrt{(\sigma_{11}\sigma_{22})}}. \tag{8}$$

The results are given in Table 11.

Table 11. *Correlation coefficients of the p_{ik}-matrices*

Author	Work	$c \times 10^2$
Shakespeare	*Othello*	−2·992
Rilke	*Cornet*	−1·673
Hesse	*Steppenwolf*	+1·259
Mann	*Buddenbrooks*	+1·033
Mann	*Zauberberg*	+0·174
Carossa	*Geheimnisse*	−1·519
Jaspers	*Der phil. Glaube*	−6·507

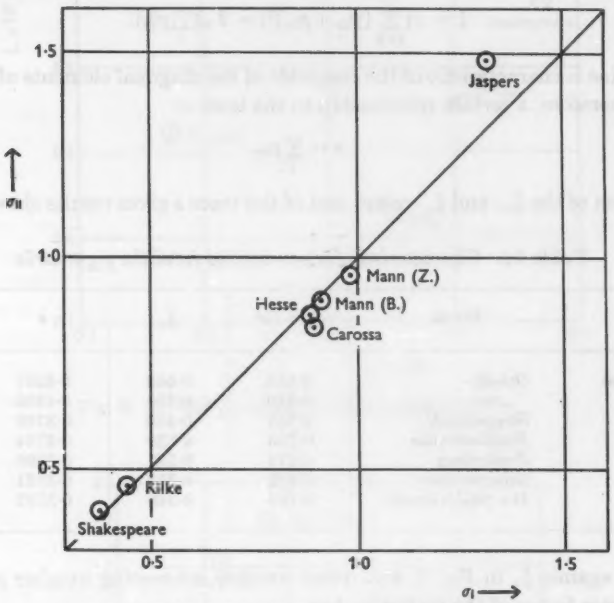


Fig. 1. Principal variances of the p_{ik}-distributions.

The sign of the correlation coefficient c is determined by the sign of σ_{12} and indicates the orientation of the correlation ellipse relative to the (i, k)-co-ordinate system. A positive sign of c means that a certain value of i is preferentially followed by a value of k, which is equal or nearly equal to i, and therefore indicates a more 'balanced' character of the text. A negative sign means that i is preferentially followed by a value of k more different from i, thus indicating a more 'rugged' structure of the text. As all our correlation ellipses are very nearly circles, it will be seen that these trends are, in our examples, very slight ones. Nevertheless, they are very interesting having regard to the nature of our problem.

2·5. *A measure for the skewness of the matrix p_{ik}*

The matrices p_{ik} of our examples (Tables 1–7) are not symmetrical ones, p_{ik} not being equal to p_{ki}. The frequency with which (i, k)-pairs are formed is not the same as the frequency with which (k, i)-pairs are formed.

We divide the p_{ik}-matrices in a well-known manner into a symmetrical and an anti-symmetrical matrix:

$$p_{ik} = \left(\frac{p_{ik}+p_{ki}}{2}\right) + \left(\frac{p_{ik}-p_{ki}}{2}\right), \tag{9}$$

and we consider the mean-square value of the deviation from the symmetry, ξ_-:

$$\xi_- = \sqrt{\{\sum_{i<k} (p_{ik}-p_{ki})^2\}}. \tag{10}$$

Correspondingly we define the values ξ_+:

$$\xi_+ = \sqrt{\{\sum_{i<k} (p_{ik}+p_{ki})^2\}} \tag{11}$$

and

$$\xi = \sqrt{\{\sum_{i=k} (p_{ik}+p_{ki})^2\}} = 2\sqrt{(\sum_i p_{ii}^2)}. \tag{12}$$

The latter value is characteristic of the ensemble of the diagonal elements of the p_{ik}-matrix and shows, therefore, a certain relationship to the trace s:

$$s = \sum_i p_{ii}. \tag{13}$$

The calculation of the ξ_-- and ξ_+-values and of the trace s gives results shown in Table 12.

Table 12. *Characteristic factors derived from the p_{ik}-matrix*

Author	Work	$\xi_- \times 10^2$	ξ_+	s	$\rho \times 10^2$
Shakespeare	*Othello*	0·585	0·668	0·6357	0·876
Rilke	*Sonnet*	0·330	0·550	0·4959	0·600
Hesse	*Steppenwolf*	0·721	0·459	0·3756	1·569
Mann	*Buddenbrooks*	0·765	0·339	0·3764	2·258
Mann	*Zauberberg*	0·274	0·357	0·3598	0·770
Carossa	*Geheimnisse*	0·422	0·372	0·3521	1·134
Jaspers	*Der phil. Glaube*	0·703	0·349	0·3193	2·014

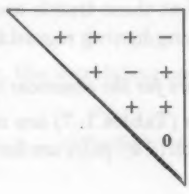
We plot ξ_- against ξ_+ in Fig. 2, and define another interesting number ρ (hence called relative skewness factor of the matrix p_{ik}) as

$$\rho = \tan\psi = \frac{\xi_-}{\xi_+}. \tag{14}$$

The values for ρ are to be found in Table 12.

We plot ρ against s in Fig. 3.

In forming the factor ξ_- the sign of $(p_{ik}-p_{ki})$ is lost. As an example we give a scheme of the signs of $(p_{ik}-p_{ki})$ of Shakespeare's *Othello*:



If we reverse the sense of the direction in which the original text is read, the i and k are reversed too and the matrix p_{ik} is transformed into the transposed matrix $\tilde{p}_{ik} = p_{ki}$, and all the signs of the scheme are reversed also.

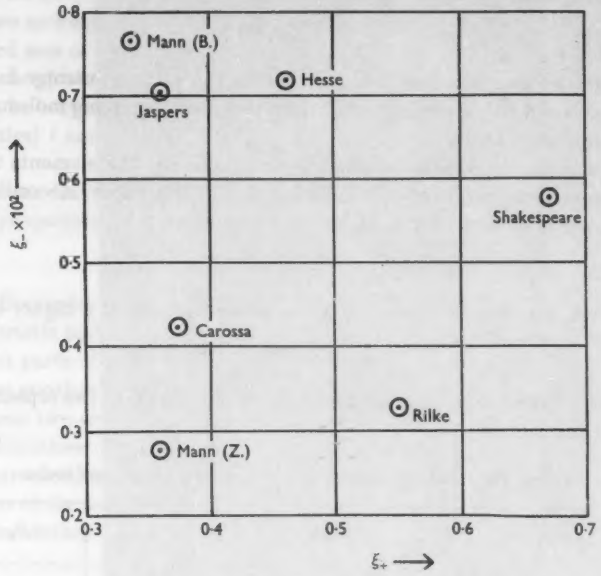


Fig. 2. ξ_-- and ξ_+-values of the p_{ik}-matrices.

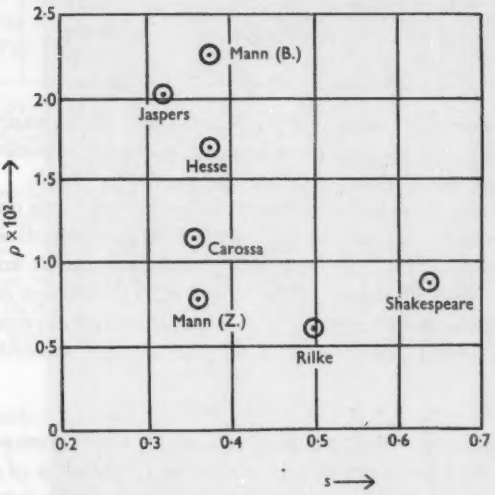


Fig. 3. ρ- against s-values of the p_{ik}-matrices.

2·6. *The entropy of a text*

We consider the text as a mixture of units. If these units are elements as studied so far, we define a mixing-entropy-factor S_1, as given already in I,

$$S_1 = -k \sum_i p_i \log p_i. \tag{15}$$

In an analogy, we can compare this factor with the mixing-entropy for a mixture of several gases, each of which is distinguished from the other by having molecules constituted of a different number of atoms.

Now, we assume the units constituting the text to be not the elements themselves but groups of elements as defined in the previous sections of this paper. Accordingly we define, for the special case of binary groups, an entropy-factor S_2:

$$S_2 = -k \sum_i \sum_k p_{ik} \log p_{ik}. \tag{16}$$

In an analogous process we construct, in the case of groups of a higher order, entropy-factors of a higher order

$$S_n = -k \sum_i \sum_k \cdots \sum_n p_{ik \ldots n} \log p_{ik \ldots n}. \tag{17}$$

For the entropy-factors S_2 we get the results shown in Table 13 (we repeat the results for S_1 from I in the same table).

Table 13. *Entropy values of the first and the second order*

Author	Work	S_1	S_2
Shakespeare	*Othello*	0·297	0·593
Rilke	*Cornet*	0·385	0·767
Hesse	*Steppenwolf*	0·499	0·974
Mann	*Buddenbrooks*	0·505	1·009
Mann	*Zauberberg*	0·516	1·030
Carossa	*Geheimnisse*	0·501	1·009
Jaspers	*Der phil. Glaube*	0·567	1·126

Table 13 shows that S_2 is very nearly equal to $2S_1$. It can easily be shown that this relation holds exactly if the formation of the (i, k)-pairs is stochastically independent and that in the case of stochastical independency generally, $S_n = n \cdot S_1$.

The definition of the S_2-matrix is not very well suited to bring out the slight (but very significant) deviations of our special p_{ik}-matrices from stochastical independency. The difference of the values S_2 for different literary works are drastic and significant, but not much more so than the differences of the values S_1. The concept of the entropy S_2 will, however, become much more interesting, if the text elements and properties of text elements are studied which lead to p_{ik}-matrices with very noticeable stochastical dependencies.

3. Fernordnung

In the previous sections the Nahordnungs relations of a text were studied. We can, theoretically, describe the relations within groups of any number n of elements by matrices $p_{ikl \ldots n}$ of the corresponding order. If we try, however, to compute, in practice, the characteristic matrices of the order three or four and to calculate the corresponding characteristic

factors, the amount of work increases enormously and therefore the method becomes, outside the Nahordnungs region, practically uninteresting. To study the question, whether at all and to what extent exact relations between distant elements or groups of elements in a text exist, we have to look for other methods.

First of all, we no longer break up the text, but we consider it as it is given by its author. Then, we regard sets of $2, 3, \ldots, n$ elements, which we assume no longer to be arranged in the immediate neighbourhood of one another; there can be (but not necessarily must be) other elements in between the $2, 3, \ldots, n$ elements of a set. The distances l for two special marks (properties) i and k, for instance, of two consecutive elements of one set may be called l_{ik}.

Now we procure the frequency distribution of the l_{ik} for the whole of the text. If the values for the properties i and k of the elements vary according to

$$1 \leqq i \leqq I, \quad 1 \leqq k \leqq I,$$

there exist I^2 frequency distributions of this sort. Starting from these distributions, we derive characteristic matrices and characteristic factors, suitable to study the relationship between distant parts of a text. For an illustration of the concept of l_{ik} cf. § 3·2.

Further, from another general point of view, we are going to study the question of possible relations between two or more elements, either part of them or all of them neighbouring or distant, using functions of influence and correlation coefficients. With the help of the latter we define a characteristic distance L that can be compared, in a physical analogy, to the mean free paths of the molecules of a mixture of chemical elements or to the range of forces between the elementary particles of a solid body or a gas.

3·1. *Frequency distributions of element distances*

We study a text with the elements e. Let the elements be characterized by marks called $i, k, \ldots$. Now consider two elements e_i and e_k, which may or may not be consecutive text elements, i and k being now special values out of the number I^2 of possible values for the (i, k)-combinations. The distance between the special elements e_i and e_k will be denoted by l_{ik}.

Now there exists obviously a frequency distribution for the distances l_{ik} defined in this special way. The number of frequency distributions of this kind amounts to I^2, as shown before.

The l_{ik}-distributions altogether comprise the p_{ik}-matrices, studied in § 2, the l_{ik}-values of which being exclusively equal to 1. In this special case we will hence describe the frequency distributions p_{ik} as ${}^1p_{ik}$.

In exactly the same way as we have studied the ${}^1p_{ik}$-matrices in the earlier sections, we can calculate the characteristic factors $(M, \xi, \rho, s, \text{etc.})$ of the matrices ${}^2p_{ik}, {}^3p_{ik}, \ldots, {}^\alpha p_{ik}$, etc., the l_{ik}-values of which are $1, 2, \ldots, \alpha$ respectively, i.e. the matrices

$${}^\alpha p_{ik} = ({}^1p_{ik}, {}^2p_{ik}, \ldots, {}^{A-1}p_{ik}),$$

A being the total number of elements of the text.

In doing this, we proceed, step by step, from the relations between the properties of consecutive elements to relations between the properties of increasingly distant elements.

In this section we have, so far, only referred to matrices of the second order. In quite an analogous way it is possible to define matrices of a higher order, say n, with which there

exist, obviously, $(n-1).(A-n+1)$ of the 'α-possibilities' just described. Therefore we have to write for these matrices of the $3, 4, \ldots, n$th order

$$^{\alpha_1, \alpha_2, \ldots \alpha_{n-1}}p_{ikl\ldots n}.$$

With the words as elements and the syllables as properties we expect

$$p_{ikl\ldots n} \to p_i \cdot p_k \cdot p_l \cdots p_n$$

for increasing α-values.

3·2. *The matrix $\bar{l}_{ik}$ of the mean distances and the matrix r_{ik} of the reciprocal values of the mean distances*

We shall now define several expressions, derived from the $^{\alpha}p_{ik}$-matrices, more interesting than the mass of the material contained in the $A-1$ matrices $^{\alpha}p_{ik}$. We begin with defining matrices of the mean distances.

To do this, the text offers various possibilities. If we denote the frequency distributions of the l_{ik} by the letter H and the corresponding distribution of the relative frequencies by

$$h(l_{ik}) = \frac{H(l_{ik})}{A-1}, \tag{18}$$

we can define a mean value $\bar{l}_{ik}$ for each combination (i, k) of properties according to

$$\bar{l}_{ik} = \sum_{l_{ik}} l_{ik}.h_{ik}, \quad \text{with} \quad h_{ik} = h(l_{ik}). \tag{19}$$

This construction, however, is obviously much too clumsy to be suited for our purposes, especially as the values of $H(l_{ik})$ do not decrease for increasing l_{ik}, so that the mean value $\bar{l}_{ik}$, as defined in (19), becomes somewhat trivial.

Therefore we define another matrix $\bar{l}_{ik}$ in the following way: Let us consider the first i-syllabled word of the text and then go along the text till we meet, for the first time, a word with k syllables. Let the distance between these two words be called l_{ik}. Then we consider the second i-syllabled word of the text and repeat the process and thus we find a second value l_{ik}. To distinguish these two l_{ik}-values, we call them $l_{ik}^{(1)}$ and $l_{ik}^{(2)}$, respectively. Proceeding in this way through the whole extent of the text we find a number, let us say z_{ik}, of l_{ik}-values. The different l_{ik}-values may be denoted generally as $l_{ik}^{(\beta)}$. We compute their mean value $\bar{l}_{ik}$

$$\bar{l}_{ik} = \frac{\sum_{\beta=1}^{z_{ik}} l_{ik}^{(\beta)}}{z_{ik}}. \tag{20}$$

This matrix describes a set of values which is more suitable for the purposes of our problem than the $\bar{l}_{ik}$ defined in (19). For obvious reasons we also consider the matrix r_{ik} of the reciprocal values of the $\bar{l}_{ik}$

$$r_{ik} = \frac{1}{\bar{l}_{ik}}. \tag{21}$$

There are other possibilities for the construction of interesting types of distance-matrices, out of which we have considered, so far, only one. Another one can be explained with the help of the following example:

Position of word	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
No. of syllables	2	5	3	1	4	7	2	1	3	2	2	4	2	5	7	6

According to the definitions given in equation (20) we calculate

$$\bar{l}_{27} = \frac{\sum_{\beta} l_{27}^{(\beta)}}{z_{27}} = \frac{5+8+5+4+2}{5} = \frac{24}{5} = 4{\cdot}8.$$

In this way the mark 7 in position 6 enters once, while the mark 7 in position 15 enters four times into our calculation. We now exclude all the l_{27}-values with the exception of the minimum values l_{27} with regard to each one of the numbers 7 in the two positions of our example; thus we obtain

$$\bar{l}'_{27} = \frac{\sum'_{\beta} l_{27}^{(\beta)}}{z'_{27}} = \frac{5+2}{2} = \frac{7}{2} = 3{\cdot}5.$$

Whenever we calculate the distance matrices or the r_{ik}-matrices in the latter way, we denote this by writing Σ' in the summations and $\bar{l}'_{ik}$ and r'_{ik} for the matrices.

In Table 14 we give an example of a r'_{ik}-matrix for Rilke's *Cornet*.

Table 14. *r'_{ik}-matrix for Rilke's* Cornet

	1	2	3	4
1	0·6231	0·9251	0·6735	0·6275
2	0·9391	0·2922	0·3212	0·2700
3	0·6774	0·3410	0·0594	0·0697
4	0·6600	0·2432	0·0729	0·0130

3·3. *Relations between the r_{ik}-matrix and characteristic factors defined before*

As pointed out in I, there exist simple relations between the $\bar{l}_{ik}$-matrices, interpreted especially for the words as elements and the numbers of syllables per word as the properties of the elements, and the relative frequency distributions of the number of syllables per word as given in the examples of I. This holds obviously for the r_{ik}-matrices too.

If we consider, for instance, the one-syllabled words only, their mean distance apart in the text is given by the length of the whole text, equal to the number $A-1$ of the sum total of the intervals between consecutive words, divided by the number A_1 of the one-syllabled words (if we neglect certain small regions at the beginning and the end of the text). Thus

$$\bar{l}_{11} = \frac{A-1}{A_1}. \tag{22}$$

Therefore, the value of r_{11} becomes

$$r_{11} = \frac{1}{\bar{l}_{11}} = \frac{A_1}{A-1} = p_1\left(1+\frac{1}{A}+\frac{1}{A^2}+\dots\right). \tag{23}$$

Under the condition $A \gg 1$, which must obviously be fulfilled, we get generally

$$r_{ii} = \frac{1}{\bar{l}_{ii}} = p_i. \tag{24}$$

The diagonal terms of the r_{ik}-matrix and also of the r'_{ik}-matrix are equal to the relative frequency distributions calculated in I. Therefore

$$\sum_i r_{ii} = \sum_i r'_{ii} = 1. \tag{25}$$

Since it is the mean values of the l_{ik} which enter into the matrices $\bar{l}_{ik}$, we may question whether or not there are relations between all the $\bar{l}_{ik}$-values and the relative frequency distributions p_i of I, i.e. beyond the very simple ones just pointed out, so that not only the $\bar{l}_{ii}$ but also the $\bar{l}_{ik}$, for which $i \neq k$, could be calculated as functions of the p_i. It is obviously possible to construct artificial types of texts, for instance, binary texts (texts with not more than two marks) with strict periodicities, which allow us to calculate all the elements of the $\bar{l}_{ik}$-matrix as functions of the relative frequencies p_i. But it is easily seen that this cannot hold generally.

The r_{ik}-matrix cannot, it is true, be interpreted as a sort of a frequency distribution in the statistical sense of the word; we can, nevertheless, study the properties of the matrix by defining factors quite analogous to the factors we defined, for instance, for the ${}^1p_{ik}$-matrix of the first part of this paper. We may also regard the r_{ik}-matrix as representing a mass distribution in a plane co-ordinate system, having i^* and k^* as the co-ordinates of the mass centre:

$$i^* = \sum_i \sum_h i r_{ih}, \quad k^* = \sum_i \sum_h k r_{ih}, \tag{26}$$

and with the moments θ_{11} and θ_{22} and the product θ_{12} of inertia as already described before.

For our example in Table 14, we find the following values:

$$i^* = 13{\cdot}893, \quad \theta_{11} = 924{\cdot}38$$
$$k^* = 13{\cdot}794, \quad \theta_{12} = 946{\cdot}61$$
$$\theta_{22} = 950{\cdot}61$$

After the well-known transformation to principal axes, we get the principal moments of inertia

$$\theta_{\mathrm{I}} = 1884{\cdot}2, \quad \theta_{\mathrm{II}} = -9{\cdot}21.$$

We further compute the values ξ_-, ξ_+, ξ and ρ in analogy to equations (10), (11), (12) and (14):

$$\xi_- = 4{\cdot}899 \times 10^{-2}, \quad \xi = 1{\cdot}382,$$
$$\xi_+ = 2{\cdot}771, \qquad \rho = 1{\cdot}768 \times 10^{-2}.$$

In analogy to S_2 defined by (16) we compute

$$S_2 = -k \sum_i \sum_h r_{ih} \log r_{ih}, \quad S_2 = 1{\cdot}6984. \tag{27}$$

It is understood, however, that this factor S_2 cannot be interpreted as simply as the factor of equation (16) in § 2·6.

3·4. *Interaction between distant parts of a text and correlation coefficients*

Consider the elements of a text in their original order. Let the position number of the elements be called x and the properties of the elements be regarded as functions $f(x)$. Now consider properties of the same sort at two different positions x_1 and x_2 of the text, i.e. the functions $f(x_1)$ and $f(x_2)$. We assume that there exists an 'influence' between the functions at different places. The influence of $f(x_1)$ on the element in the place x_k will be formulated with the help of an influence function $\phi(x_1, x_k)$, as follows:

$$\phi(x_1; x_k) f(x_1).$$

Similarly, we expect an influence from elements in other places $x_2, \ldots, x_i$ of the text on the element in the position x_k. If the sum total of all these influences (and nothing else) determines the value of $f(x_k)$, we may write

$$f(x_k) = \sum_{i \neq k} \phi(x_i; x_k) f(x_i), \tag{28}$$

which, in a calculus with continuous variables, amounts to an integral equation.

We conclude that our characteristic numbers concerning Nahordnung and, still more, Fernordnung will turn out to be of decisive value for the exact description of the peculiarities of the style of a special literary work, if we can expect that a certain property of the text elements at a place x_i influences the properties of elements at distant places.

If such influences between distant text elements exist, we may speak, in the realm of statistics, of correlative relations, which can be formulated exactly by correlation coefficients. One of these has been made use of in § 2·4. It is based on the moments of the second order of a frequency distribution:

$$c = \frac{\sigma_{12}}{\sqrt{(\sigma_{11}\sigma_{22})}}.$$

3·5. *Application of correlation coefficients on literary texts and the definition of the characteristic distance L*

Consider, as in the previous section, a text element in the position i with the mark f_i and a text element in the position $k = i + l$ with mark f_k and form the correlation coefficient q

$$q_l = \frac{\overline{f_i \cdot f_{i+l}}}{\sqrt{(\overline{f_i^2} \cdot \overline{f_{i+l}^2})}}, \tag{29}$$

where

$$\overline{f_i f_{i+l}} = \frac{1}{A-l} \sum_{\gamma=1}^{A-l} f_i^{(\gamma)} f_{i+l}^{(\gamma)}, \quad \overline{f_i^2} = \frac{1}{A-l} \sum_{\gamma=1}^{A-l} (f_i^{(\gamma)})^2 \tag{30}$$

(with A = the number of elements of the whole text). The correlation coefficient q_l is a function of the value of l.

Now define a characteristic distance by

$$L = \frac{1}{A-1} \sum_{l=1}^{A-1} q_l. \tag{31}$$

If there exist binding forces (influences) within the structure of a text, according to our concepts of Nahordnung and Fernordnung, we can make a distinction between short-range and long-range forces, comparable to the cohesion forces and the electric forces in physics, respectively. The value of the characteristic length L represents an over-all measure of the range of the binding forces of the whole of a text. This factor L, therefore, gives the exact answer to one of the most interesting questions of our problem, namely, whether there exist interactions (binding forces) within the structure of a given text, and if they do exist, of what range they are exactly.

As pointed out before, it would not be of much interest to compute correlation coefficients or characteristic distances L with respect to properties of text elements, as to which we cannot expect any noticeable interaction between distant parts of the text. If, for instance, we should choose, in a prose text, the words as elements and their syllable numbers as their

properties, the q_l-values could not be expected to bring out especially characteristic distinctions about the style of different authors. In this case the value of the distance L would turn out to be invariably in the neighbourhood of unity. It is easily seen that useful results can be expected only for elements and properties with respect to which a fairly strong influence between distant parts of the text is to be expected, i.e. for instance, for metric elements in a text, written in verse, or for a text written in prose, but (stochastically) interspersed with noticeable parts of poetic language.

It will be shown, however, that the definition of style characteristics, especially suitable for texts of these types, requires a new approach. Therefore, non-trivial examples of correlation coefficients defined as in equation (29) and the characteristic length L will be given in another communication.

REFERENCE

Fucks, W. (1952). On mathematical analysis of style. *Biometrika*, **39**, 122.

A DISTRIBUTION-FREE *k*-SAMPLE TEST AGAINST ORDERED ALTERNATIVES

By A. R. JONCKHEERE

Department of Psychology, University College, London

1. Introduction

An experimenter is frequently able to state, on the basis of experience or theory, the expected *rank order* of magnitude of the effects of different experimental treatments. Suppose, for example, that an experiment is performed to determine the effect of different degrees of stress upon the performance of some task of manual dexterity. The data for analysis might consist of the number of errors made by four different groups of subjects, each working under one of the conditions: high, medium, low, or minimal stress. The experimenter hopes to refute the hypothesis that the four samples of data could be considered as randomly drawn from the same population. However, experience or theory lead him to expect that the number of errors made increases with the degree of stress (his alternative hypothesis). Hence he desires a test of significance especially sensitive to those differences which, while tending to reject the null hypothesis that the samples are all drawn from the same population, at the same time lends support to the specified alternative in question.

The customary one-way analysis of variance does not satisfy this demand: the F-ratio is independent of the order in which the group means occur. Furthermore, attempts to combine the probabilities yielded by an F-ratio and some coefficient of rank correlation between the expected and the obtained order of the group means are faced with a number of difficulties. In the first place, it is necessary to combine a continuous and a discrete probability distribution, a procedure requiring special consideration and some arduous computation for each case as it arises (see, for example, F. N. David (1947)). In the second place, a complication arises owing to the fact that, because of the essential discontinuity of the measuring scales we have to employ in practice, the F-ratio may have a determinable value even when several of the group means are identical. Should this occur, the appropriate sampling distribution of the rank correlation between the predicted and the obtained rank order of means becomes troublesome. A similar difficulty arises when attempts are made to combine the probabilities yielded by the χ^2 criterion and the rank correlation of the ratios involved in a $2 \times k$ contingency table—as suggested by Hotelling & Pabst (1936). In this latter case, if the possibility of ties in the values of the ratios in a $2 \times k$ contingency table is excluded, then the values of χ^2 and the rank correlation are no longer independent for small samples. The difficulty in these situations is that while we can fairly easily determine the distribution of rank correlations for a *fixed* number and extent of ties, what seems to be required is the distribution when any manner of ties may occur, including the case where there are none. For unless there were other evidence at hand, it would be unreasonable to suppose that future repetitions of the same experiment would lead in every instance to exactly the same number and extent of ties. This complication, whatever may be the other merits of 'conditional' statistics, goes in fact farther than the present context, and is believed

to render difficult any rational account of the significance tests commonly applied when ties occur in rankings*.

A further aspect of many experiments is that there appears to be no suitable metric by which the different treatments to be investigated may be characterized. All that an experimenter is able to assert, particularly in psychology, is that they may be ranked in some order, such as increasing 'stress'. When this is the case, it is no longer possible to use any form of regression analysis, since the independent variate is not suitably quantified.

We have then the problem of assessing whether the results obtained from a one-way analysis of variance design are 'significant' and in conformity with a hypothesis as to the rank order for the size of the effects yielded by different treatments. When there are only two treatments under consideration, the problem is recognized and treated in terms of the usual distinction made between single- and double-tailed tests of significance. If the experimenter is indifferent as to which of the two treatments will give the larger results, a double-tailed test is employed. If, on the other hand, it is expected that a particular treatment will be more effective than the other, an appropriate single-tailed test is used. The proposed test attempts to extend this distinction to situations in which there are more than two groups by specifying one out of the many alternatives to the null hypothesis.

Such a test has obvious applications in the study of time series. Suppose, for instance, that on each of n successive occasions, any one of k events may occur. Then we can test the hypothesis that the k events occur randomly in the series of n occasions, against the alternative that they tend to occur in a particular ordered time sequence. For example, the births of siblings might be of three types: normal, abnormal in respect of some characteristic, and stillbirths. The effect of birthrank on the type of birth could be tested here when, for instance, the alternative to randomness is that the earlier births tend to be normal, the later births abnormal, with finally the appearance of stillbirths. The application here is an extension of the method employed by Haldane & Smith (1948) for a similar situation, but where there are only two alternatives for characterizing the successive births.

2. The test proposed

Let $(X_{11}, X_{12}, \ldots, X_{1m_1}), \ldots, (X_{i1}, \ldots, X_{ia_i}, \ldots, X_{im_i}), \ldots, (X_{k1}, \ldots, X_{km_k})$ be k samples of size $m_1, m_2, \ldots, m_i, \ldots, m_k$, randomly drawn from populations with continuous cumulative distributions $F_1(X), F_2(X), \ldots, F_i(X), \ldots, F_k(X)$ respectively, and arranged such that the first suffix of the X's is in the order implied by the alternative hypothesis

$$F_1(X) < F_2(X) < \ldots < F_i(X) < \ldots < F_k(X) \quad \text{for all } X.$$

Generally, if X_{ia_i} is the a_ith value in the ith sample drawn from a population with c.d.f. $F_i(X)$, we wish to test the hypothesis that $F_i(X) = F_j(X)$, $(i, j = 1, \ldots, k;\ i \neq j)$, against the alternative that $F_i(X) < F_j(X)$ $(i < j)$ for all X.

Let

$$p_{ia_i ja_j} = \begin{cases} 1 & \text{if} \quad X_{ia_i} < X_{ja_j}, \\ 0 & \text{if} \quad X_{ia_i} > X_{ja_j}, \end{cases}$$

where $\quad i = 1, \ldots, k-1; \quad j = 1+i; \quad a_i = 1, \ldots, m_i; \quad a_j = 1, \ldots, m_j$.†

* Though the test to be presented involves the distribution of a rank correlation with ties, the above-mentioned difficulties do not apply. The number and extent of ties is controlled by the number and sizes of samples, characteristics which are fairly readily reproducible in a sequence of experiments.

† See note added on p. 143 below.

Let

$$p_{ij} = \sum_{a_i=1}^{m_i} \sum_{a_j=1}^{m_j} p_{ia_i ja_j}$$

and finally, let

$$S = 2 \sum_{i=1}^{k-1} \sum_{j=1+i}^{k} p_{ij} - \sum_{i=1}^{k-1} \sum_{j=1+i}^{k} m_i m_j. \qquad (1)$$

It is the statistic S that we propose to use for testing the null hypothesis against the alternative of ordered cumulative distribution functions.*

A short numerical example may make the method of computation of S clearer while at the same time showing the use of the test. Suppose that the following values are obtained in an experiment involving four independent groups:

	I	II	III	IV
	19	21	40	49
	20	61	99	110
	60	80	100	151
	130	129	149	160
Mean	57·25	72·75	97·00	117·50

The experimenter wishes to test the hypothesis that the four samples have come from the same population against the alternative that the populations are such that the values from the samples I, II, III, IV are in an expected order of increasing value. It will be noted that the arithmetic means are in the predicted order. For the computation of S we have

$$m_1 = m_2 = m_3 = m_4 = 4, \quad k = 4, \quad p_{12} = 11, \quad p_{13} = 12, \quad p_{14} = 13,$$
$$p_{23} = 11, \quad p_{24} = 12, \quad p_{34} = 12.$$

Hence, using (1), $S = 2 \times 71 - 96 = 46$.

It will be seen that the computation is quite simple. The samples are ranged in the order implied by the alternative hypothesis under consideration. Then, for each sample in turn, we determine for each value the number of items which are larger in all the succeeding samples. This gives sums of the values of p_{ij}. Employing formula (1), we can then readily obtain the appropriate value of S.

The value of S yielded by the above computation is the same as that given by the ordinary procedure for the calculation of Kendall's S between two rankings, when one of them contains ties (Kendall, 1948). In fact, dividing the value of S by the maximum value it can have (when ties are taken into account) gives the usual τ coefficient of rank correlation. Here

$$\tau = \frac{S}{\sum_{i=1}^{k-1} \sum_{j=1+i}^{k} m_i m_j},$$

and with the example given, $\tau = \frac{46}{96} = 0{\cdot}48$. In the present test the closer τ is to unity, the more the sample data are in conformity with the ordered hypothesis alternative to that of randomness.

Employing the probability tables given later, we find $\Pr[S \geqslant 46] = 0{\cdot}0168$, and thus the experimenter would with some confidence reject the null hypothesis, and accept an alternative that the sample came from populations which were stochastically ordered in the series I, II, III, IV.

* If $F_i(X) < F_j(X)$ for all X, then it is obvious that if M_i and M_j are the medians of the distributions, $M_i < M_j$. On the other hand, the shapes of the two distributions may be identical. It is believed that the present test is more sensitive to differences of location between the population distributions than to differences of scale or skewness.

It is interesting to observe that if an analysis of variance is performed on the data of this example we obtain, as a between-to-within-groups ratio, $F_{3,12} = 1{\cdot}216$ and

$$\Pr[F_{3,12} \geqslant 1{\cdot}216] = 0{\cdot}346.$$

Hence the experimenter would have accepted with some confidence the null hypothesis that the four samples came from populations having the same means. The difference in the significance levels yielded by the two tests is, of course, explained by the greater variety of alternatives to the null hypothesis allowed for by analysis of variance techniques.

3. Previous work

The test just described is a simple extension to several groups of a procedure which has been frequently recommended for the problem of deciding whether two samples have come from the same population. Wilcoxon (1945) originally gave a form of this test, where the statistic to be computed was the sum of ranks of the items in one of the two samples, and supplied a few probability levels. When more than two samples are considered there appears to be no simple method for computing the value of S in terms of the sums of the ranks of the items in each sample. Festinger (1946) provided tables for samples of different size, but the asymptotic distribution was investigated only later by Mann & Whitney (1947), who besides showing that the normal distribution is the limiting form when *both* sample sizes increase without limit, also showed that the test was consistent with respect to the one-sided alternative $F_1(X) < F_2(X)$. By investigating the differential of the power function of the test, van der Vaart (1950) was able to show that the test was unbiased against a one-sided alternative. The power of the test has been recently discussed by Sundrum (1953). In a different context Haldane & Smith (1948) employed the same test for the detection of birth-order effects, giving the general expression for the cumulants of the statistic.

An attempt to extend the test to more than two groups has been made by Whitney (1951), who considered the distribution of a statistic related to S for the case of three groups. Unfortunately, by the use of his test, there is a large amount of information lost concerning the interrelations of the ranks of the items between the three groups, which is relevant to accepting the hypothesis alternative to randomness that the three populations are stochastically ordered in increasing value.

There have of course been many attempts to devise tests based on ranks which are in some sense analogous to the usual one-way analysis of variance procedures. Kruskal & Wallis (1952) give an account of these methods, but in so far as they do not specifically take as alternative to the null hypothesis, a hypothesis that the populations from which the samples are drawn are stochastically ordered, they differ from the test proposed here, and their limit distributions tend to be of Type III form.

K. Iyer (1952), in the course of two articles, briefly developed a k-sample test employing essentially the same statistic as S, but did not discuss the situation for which the test seems to be most suitable. As recommended by Iyer, the test does not appear to be very sensitive against the general alternatives he apparently had in mind.

4. The sampling distribution of S

In view of the identity of the statistic employed in this test and Kendall's S coefficient of correlation rank when one of the ranks contains ties, the derivation of the appropriate sampling distribution will be carried out in terms of ranks.

Consider n ranked items such that there are k sets of ties, of extent m_r ($r = 1, 2, \ldots, k$) and $\sum_{r=1}^{k} m_r = n$. Then according to the procedure employed in the previous section for computing the value of S between these tied and n untied ranks, a particular set of ties of extent m_r will make no contribution to the total value of S, since no internal comparisons are made. If, however, they were untied, their contribution to the total would be equivalent to the value of S yielded by an ordinary untied ranking of m_r items.

Let $f(n)$ stand for the probability-generating function of the values of S when both rankings consist of n untied items. Then, from the above reasoning, it follows that

$$f(n) = \prod_{r=1}^{k} f(m_r) f(m_1, m_2, \ldots, m_k),$$

where $f(m_1, m_2, \ldots, m_k)$ stands for the probability-generating function of values of S when one of the rankings consists of n untied items and the other contains ties as in the general case mentioned above.

Hence we have

$$f(m_1, m_2, \ldots, m_k) = \frac{f(n)}{\prod_{r=1}^{k} f(m_r)}. \tag{2}$$

Now Kendall has shown (1948) that in a universe of equiprobable permutations of ranks,

$$f(n) = (n!)^{-1} \prod_{s=1}^{n} (X^{-(s-1)} + X^{-(s-3)} + \ldots + X^{(s-3)} + X^{(s-1)}).$$

Making the substitution in (2) we have

$$f(m_1, \ldots, m_k) = \frac{(n!)^{-1} \prod_{s=1}^{n} (X^{-(s-1)} + X^{-(s-3)} + \ldots + X^{(s-3)} + X^{(s-1)})}{\prod_{r=1}^{k} \left\{ (m_r!)^{-1} \prod_{s=1}^{m_r} (X^{-(s-1)} + X^{-(s-3)} + \ldots + X^{(s-3)} + X^{(s-1)}) \right\}}. \tag{3}$$

Putting $X = e^{it}$ in this last expression, and employing substitutions made by Silverstone (1950) and Moran (1950) in a similar type of derivation, gives the characteristic function

$$\phi(t) = \frac{(n!)^{-1} (\sin t)^{-n} \prod_{s=1}^{n} \sin (st)}{\prod_{r=1}^{k} \left\{ (m_r!)^{-1} (\sin t)^{-m_r} \prod_{s=1}^{m_r} \sin (st) \right\}},$$

and, making use of the relation

$$\log \sin x - \log x = -\sum_{\alpha=1}^{\infty} \left\{ \frac{x^{2\alpha} B_\alpha 2^{2\alpha-1}}{\alpha (2\alpha)!} \right\},$$

we get as the cumulant-generating function

$$K(t) = \log \phi(t) = \sum_{\alpha=1}^{\infty} \frac{B_\alpha 2^{2\alpha-1}}{\alpha (2\alpha)!} t^{2\alpha} \left(n - \sum_{p=1}^{n} p^{2\alpha} \right) - \sum_{r=1}^{k} \left\{ \sum_{\alpha=1}^{\infty} \frac{B_\alpha 2^{2\alpha-1}}{\alpha (2\alpha)!} t^{2\alpha} \left(m_r - \sum_{p=1}^{m_r} p^{2\alpha} \right) \right\},$$

where B_α is the αth Bernoulli number ($B_1 = \frac{1}{6}$, $B_2 = \frac{1}{30}$, ...). Now κ_s, the sth cumulant of the distribution of S, is given by the coefficient of $(s!)^{-1} (it)^s$ in this series. Hence all the odd cumulants and moments are zero, while

$$\kappa_{2\alpha}(m_1, \ldots, m_k) = (-)^{\alpha-1} \frac{B_\alpha 2^{2\alpha-1}}{\alpha} \left\{ \sum_{p=1}^{n} p^{2\alpha} - \sum_{r=1}^{k} \sum_{p=1}^{m_r} p^{2\alpha} \right\}, \tag{4}$$

where $\kappa_2 (m_1, m_2, \ldots, m_k)$ stands for the general even-order cumulant of the distribution of S when one of the arrays of ranks contains sets of ties of any possible extent and number.

5. The cumulants of S

The first four cumulants may be obtained by expanding (4) and become

$$\kappa_1(m_1, \ldots, m_k) = 0,$$

$$\kappa_2(m_1, \ldots, m_k) = \frac{1}{18}\left\{n^2(2n+3) - \sum_{r=1}^{k} m_r^2(2m_r+3)\right\},$$

$$\kappa_3(m_1, \ldots, m_k) = 0,$$

$$\kappa_4(m_1, \ldots, m_k) = -\frac{1}{225}\left\{n^3(6n^2+15n+10) - \sum_{r=1}^{k} m_r^3(6m_r^2+15m_r+10)\right\},$$

and hence $\gamma_1 = 0,$

$$\gamma_2 = -\frac{36}{25}\frac{\left\{n^3(6n^2+15n+10) - \sum_{r=1}^{k} m_r^3(6m_r^2+15m_r+10)\right\}}{\left\{n^2(2n+3) - \sum_{r=1}^{k} m_r^2(2m_r+3)\right\}^2},$$

where $$\sum_{r=1}^{k} m_r = n.$$

If k is a factor of n, so that $km = n$, then the groups are of equal size, m, and we have

$$\kappa_1(m) = 0,$$

$$\kappa_2(m) = \tfrac{1}{18}\{k(k-1)\,m^2(2m(k+1)+3)\},$$

$$\kappa_3(m) = 0,$$

$$\kappa_4(m) = -\tfrac{1}{225}\,k(k-1)\,m^3\{k^3(6m^2) + k^2(6m^2+15m) + k(6m^2+15m+10) + 6m^2+15m+10\},$$

and hence

$$\gamma_1 = 0,$$

$$\gamma_2 = -\frac{36}{25}\frac{\{k^3(6m^2) + k^2(6m^2+15m) + k(6m^2+15m+10) + 6m^2+15m+10\}}{(k-1)\,km\{2m(k+1)+3\}^2}.$$

If $\kappa_{2\alpha}(n)$ stands for 2αth cumulant of S for the usual untied rankings of n items, we may express (4) as

$$\kappa_{2\alpha}(m_1, m_2, \ldots, m_k) = \kappa_{2\alpha}(n) - \sum_{r=1}^{k} \kappa_{2\alpha}(m_r),$$

where $\sum_{r=1}^{k} m_r = n$, a result obtained by Iyer (1952) employing a more cumbrous method.

If in (4) we put $m_1 = m_2 = \ldots = m_k = 1$, then $k = n$, and we have

$$\kappa_{2\alpha}(n) = (-1)^{\alpha-1}\frac{B_\alpha\, 2^{2\alpha-1}}{\alpha}\left\{\sum_{p=1}^{n} p^{2\alpha} - n\right\},$$

an expression obtained previously by Silverstone (1950) and Moran (1950). When $m_1 = h$ and $m_2 = k-h$, $m_3 = m_4 = \ldots = m_k = 0$, we get the result obtained by Haldane & Smith (1948) for the even-order cumulants of a distribution studied in connexion with birth-order effects, in which their symbol A is replacable by $\frac{1}{2}\{h(k+1) - S\}$ when $2h > k$.

When $m_1 = m$ and $m_2 = n$, $m_3 = m_4 = \ldots = m_k = 0$, we get, after suitable reduction, the expressions given by Mann & Whitney (1947) for the first four moments of the distribution of U, a statistic studied in connexion with a two-sample test, and which is related to the coefficient S by $S = m_1 m_2 - 2U$.

If α is put equal to 1, we get the expression for the variance of the distribution of S in the most general case when one of the rankings contains ties of any extent and number, as given by Kendall (1948).

6. THE EXTREME VALUES OF THE CUMULANTS

It is useful both for practical and theoretical reasons to know what are the maximum and minimum values which the cumulants may take for various partitions of the total sample of n items among k groups. Formally we require the maxima and minima of

$$\kappa_{2\alpha}(m_1, \ldots, m_k) = (-)^{\alpha-1}\frac{B_\alpha 2^{2\alpha-1}}{\alpha}\left\{\sum_{p=1}^{n} p^{2\alpha} - \sum_{r=1}^{k}\sum_{p=1}^{m_r} p^{2\alpha}\right\},$$

where the m_r's may vary subject to the side conditions that $\sum_{r=1}^{k} m_r = n$, and that m_r may only take integer values greater than zero. Further, since the values of α and n are taken as constant, we may simply consider the extreme values of

$$\sum_{r=1}^{k}\sum_{p=1}^{m_r} p^{2\alpha}, \quad \text{where} \quad \sum_{r=1}^{k} m_r = n.$$

Consider now the partition of n such that

$$m_1 = m_2 = \ldots = m_{k-1} = 1 \quad \text{and} \quad m_k = n-k+1. \tag{5}$$

Then
$$\sum_{r=1}^{k}\sum_{p=1}^{m_r} p^{2\alpha} = \sum_{p=1}^{n-k+1} p^{2\alpha} + (k-1) \equiv G_1(m_r), \quad \text{say.}$$

Any other possible partition may be obtained from (5) by putting $m_1 = 1+d_1$, $m_2 = 1+d_2$, $\ldots$, $m_{k-1} = 1+d_{k-1}$, and $m_k = n-k+1-d$, where $\sum_{r=1}^{k-1} d_r = d$ and the d's are so arranged that $0 \leqslant d_1 \leqslant d_2 \leqslant \ldots \leqslant d_{k-1} \leqslant n-k-d$.

With this latter partition we have

$$\sum_{r=1}^{k}\sum_{p=1}^{m_r} p^{2\alpha} = \sum_{p=1}^{n-k+1-d} p^{2\alpha} + \sum_{r=1}^{k-1}\sum_{p=1}^{1+d_r} p^{2\alpha} \equiv G_2(m_r), \quad \text{say.}$$

Then
$$G_1(m_r) - G_2(m_r) = \sum_{p=1}^{n-k+1} p^{2\alpha} + (k-1) - \sum_{p=1}^{n-k+1-d} p^{2\alpha} - \sum_{r=1}^{k-1}\sum_{p=1}^{1+d_r} p^{2\alpha}$$

$$= \sum_{p=n-k+1-d+1}^{n-k+1} p^{2\alpha} - \sum_{r=1}^{k-1}\sum_{p=2}^{1+d_r} p^{2\alpha}$$

$$= \sum_{r=1}^{k-1}\sum_{s=1}^{d_r}\left\{\left(n-k+1-d+\sum_{t=1}^{r-1} d_t + s\right)^{2\alpha} - (1+s)^{2\alpha}\right\},$$

after some manipulation.

But in view of the inequalities given above, the expression within the brackets will always be positive or zero, and hence the maximum value of $\sum_{r=1}^{k}\sum_{p=1}^{m_r} p^{2\alpha}$ is given by $\sum_{p=1}^{n-k+1} p^{2\alpha} + (k-1)$.

By a similar form of argument, it may be shown that if $n = Ak+B$, $(0 \leqslant B < k)$, then the partition of the groups given by

$$m_1 = m_2 = \ldots = m_{k-B} = A, \quad m_{k-B+1} = \ldots = m_k = A+1, \tag{6}$$

yields a minimum value of $\sum_{r=1}^{k}\sum_{p=1}^{m_r} p^{2\alpha}$ equal to $(k-B)\sum_{p=1}^{A} p^{2\alpha} + B\sum_{p=1}^{A+1} p^{2\alpha}$.

Hence the maximum value of $\kappa_{2\alpha}$ is given by the partition (6) and is equal to

$$\max \kappa_{2\alpha} = (-)^{\alpha-1}\frac{B_\alpha 2^{2\alpha-1}}{\alpha}\left\{\sum_{p=1}^{n} p^{2\alpha} - (k-B)\sum_{p=1}^{A} p^{2\alpha} - B\sum_{p=1}^{A+1} p^{2\alpha}\right\},$$

while the minimum value of $\kappa_{2\alpha}$ is given by the partition (5) and is equal to

$$\min \kappa_{2\alpha} = (-)^{\alpha-1}\frac{B_\alpha 2^{2\alpha-1}}{\alpha}\left\{\sum_{p=1}^{n} p^{2\alpha} - \sum_{p=1}^{n-k+1} p^{2\alpha} - (k-1)\right\}.$$

In particular, the variance of the distribution of S is a maximum when the ties are as nearly equal in size as possible.

7. The large sample distribution

The standardized αth cumulant of the distribution of S is

$$\gamma_\alpha = \frac{\kappa_{2\alpha}}{(\kappa_2)^\alpha} = \frac{(-)^{\alpha-1}B_\alpha 2^{2\alpha-1}\left\{\sum_{p=1}^{n} p^{2\alpha} - \sum_{r=1}^{k}\sum_{p=1}^{m_r} p^{2\alpha}\right\}}{\alpha B_1^\alpha\left\{\sum_{p=1}^{n} p^{2} - \sum_{r=1}^{k}\sum_{p=1}^{m_r} p^{2}\right\}^\alpha}.$$

Now
$$\sum_{p=1}^{n} p^{2\alpha} = n^{2\alpha+1}\left\{\frac{1}{(2\alpha+1)} + \frac{1}{2n} + \sum_{i=1}^{\alpha}\frac{(-)^{i-1}B_i(2\alpha)!}{(2i)!\{2(\alpha-i)+1\}!}\frac{1}{n^{2i}}\right\},$$

Hence, using this expression, we have

$$\gamma_\alpha = \frac{(-)^{\alpha-1}B_\alpha 2^{2\alpha-1}\left\{\frac{1}{(2\alpha+1)}u(2\alpha+1) + \frac{1}{2n}u(2\alpha) + \sum_{i=1}^{\alpha-1}\frac{(-)^{i-1}B_i(2\alpha)!}{(2i)!\{2(\alpha-i)+1\}!}\frac{u\{2(\alpha-i)+1\}}{n^{2i}}\right\}}{B_1^\alpha \alpha n^{\alpha-1}\left\{\frac{1}{3}u(3) + \frac{1}{2n}u(2)\right\}^\alpha}, \quad (7)$$

where
$$u(x) = 1 - \sum_{r=1}^{k}\left(\frac{m_r}{n}\right)^x.$$

Taking k, the number of groups, as finite, let the size of the groups increase with n so that $\lim_{n\to\infty}\frac{m_i}{n} = a_i$. Further, let $a_i = A_i$ for $i = 1, \ldots, t$ and $a_i = 0$, for $i = t+1, \ldots, k$. Now

$$\sum_{i=1}^{t}\left(\frac{m_i}{n}\right) + \sum_{i=t+1}^{k}\left(\frac{m_i}{n}\right) = 1,$$

and proceeding to the limit we have $\sum_{i=1}^{t} A_i = 1$. Hence

$$\lim_{n\to\infty}\sum_{r=1}^{k}\left(\frac{m_r}{n}\right)^{2\alpha+1} = \sum_{i=1}^{t} A_i^{2\alpha+1} \leqslant \left[\sum_{i=1}^{t} A_i\right]^{2\alpha+1} = 1.$$

If $t > 1$, then $\lim_{n\to\infty}\sum_{r=1}^{k}\left(\frac{m_r}{n}\right)^{2\alpha+1} < 1$ and $\lim_{n\to\infty} u(2\alpha+1) > 0$, $(\alpha > 0)$. Hence, from (7), it can be seen that $\lim_{n\to\infty}\gamma_\alpha = 0$. Hence if *more than one group is allowed to increase without limit as* $n\to\infty$, then the limit distribution of S will be normal. If, on the other hand, $t = 1$, then $\lim_{n\to\infty}\sum_{r=1}^{k}\left(\frac{m_r}{n}\right)^{2\alpha+1} = 1$. In this case the groups will be such that $m_1 = n-d$ and $\sum_{i=2}^{k} m_i = d$, where d is constant with respect to n.

Further, we have now

$$u(x) = 1 - \sum_{r=1}^{k} \left(\frac{m_r}{n}\right)^x = \frac{1}{n}\left\{xd + O\left(\frac{1}{n}\right) - \frac{1}{n^{x-1}} \sum_{r=2}^{k} m_r^x\right\},$$

hence $\lim_{n\to\infty} nu(x) = xd$, and, taking the limit of (7), we obtain

$$\lim_{n\to\infty} \gamma_\alpha = \frac{(-)^{\alpha-1} B_\alpha 2^{\alpha-1}}{B_1^\alpha \alpha d^{\alpha-1}}.$$

Hence *if only one group* is allowed to increase without limit as $n \to \infty$, then the limit distribution of S will be non-normal and platykurtic. However, with a large number of groups, or with a large number of items in the groups, the value of d should be large enough to make the distribution normal for most practical purposes. We have, in particular,

$$\lim_{n\to\infty} \gamma_2 = -\frac{1{\cdot}2}{\sum_{i=2}^{k} m_i} = -\frac{1{\cdot}2}{\overline{m}(k-1)},$$

where $\overline{m}$ stands for the mean size of the $k-1$ constant groups. In the extreme case where $\overline{m} = 1$, we have

$$\lim_{n\to\infty} \gamma_2 = -1{\cdot}2/(k-1).$$

It follows that for large samples we may in most cases assume

$$\frac{S}{\sqrt{\left[\frac{1}{18}\left\{n^2(2n+3) - \sum_{r=1}^{k} m_r^2(2m_r+3)\right\}\right]}}$$

to have a standard normal distribution. Since the interval between all possible adjacent values of S is always two, an improvement in the approximation to the true distribution will be obtained if unity is subtracted from an obtained value of S, prior to its division by the standard deviation.

A somewhat closer approximation to the true distribution of S may be obtained by fitting a Pearson Type II curve and transforming to a Type VII or t-distribution. This is suggested by the fact that the range of S is finite, the distribution is symmetrical, and the value of β_2 less than 3.

Representing the distribution by a Type II curve we have

$$f = \frac{1}{aB(m+1, \frac{1}{2})}\left(1 - \frac{S^2}{a^2}\right)^m dS \quad (-a \leqslant S \leqslant +a),$$

where

$$m = \frac{9 - 5\beta_2}{2\beta_2 - 6}, \quad a = \sqrt{\frac{-2\mu_2\beta_2}{\beta_2 - 3}}.$$

If now we put

$$t^2 = \frac{S^2}{a^2}\frac{2(m+1)}{(1 - S^2/a^2)} \quad (-\infty \leqslant t \leqslant +\infty),$$

we have

$$f = \frac{1}{B(m+1, \frac{1}{2})\sqrt{(2m+2)}} \frac{dt}{(1 + t^2/(2m+2))^{\frac{1}{2}(2m+3)}},$$

which is the t-distribution with ν degrees of freedom, where $\nu = 2(m+1)$.

Hence the distribution of S will be approximated by

$$t_\nu = S\sqrt{\frac{\nu}{(\nu+1)\mu_2 - S^2}},$$

distributed as 'Students' t with ν degrees of freedom, where

$$\nu = -3(2+\gamma_2)/\gamma_2.$$

Considering the case where, as $n \to \infty$, we have $\gamma_2 \to -0$; then $\nu \to \infty$ and $t_\nu \to S/\sqrt{\mu_2}$, giving the normal approximation obtained above. A continuity correction can again be applied here by deducting unity from the value of S prior to computing t_ν.

8. The exact distribution of S for small samples

With small samples we may always obtain the exact sampling distribution of S by use of the general probability-generating function (2). The task of tabling all the possible combinations of number of samples and sample sizes is, however, formidable, even for small values of n. We give therefore tables only for situations where the sizes of the samples are all equal.* In computations for this purpose a number of recursive formulae can be employed for reducing the labour involved in determining the coefficient of X.

We have

$$f_k(m) = f_k(m-1)\frac{[(km-k)!]\,m^k}{[(km)!]}X^{-\frac{1}{2}k(k-1)(2m-1)}\frac{\prod_{s=k(n-1)+1}^{km}(1-X^{2s})}{(1-X^{2m})^k},$$

where $f_k(m) \equiv f(m, \ldots, m)$ (m repeated k times).

Now $f_k(1)$ is the distribution of the values of S when each sample only contains one number and has already been tabled by Kendall (1948). Hence by the formula given above, tables of increasing equal sample sizes for a fixed k may be fairly easily computed using the method of detached coefficients. Table 3 gives the probability of obtaining values of S greater than or equal to a fixed value for all cases of equal-sized samples where the normal distribution does not appear to be adequate.

When the samples are of unequal size and small it is necessary to compute the probability distribution of S from the general generating function (2). A more convenient form is as follows:

$$f(m_1, m_2, \ldots, m_k) = \frac{(n!)^{-1}X^{-\frac{1}{2}\left(n^2-\sum_{r=1}^{k}m_r^2\right)}\prod_{s=1}^{n}(1-X^{2s})}{\prod_{r=1}^{k}\left\{(m_r!)^{-1}\prod_{s=1}^{m_r}(1-X^{2s})\right\}}.$$

Expanding this expression by the methods of detached coefficients we obtain the exact probability distribution of S for any set of m_i's.

It is interesting to observe that as the sample sizes become more and more heterogeneous, so the distribution of S becomes more and more platykurtic. Table 1 shows the values of γ_2 corresponding to the partitions of $n = 10$ into five samples. The tendency shown by these figures would seem to be in general accord with the result obtained above; that the limiting distribution is non-normal when at most one of the groups is allowed to increase without limit.

The accuracy of the normal and t-distribution approximations is shown in Table 2. This table gives the exact probability distribution of S for three samples of sizes 10, 1 and 1, together with the normal and t-distribution approximations. The value of γ_2 is in this case $-0{\cdot}6137$; the variance is 87·6667, and the appropriate equivalent degrees of freedom for

* Sillito (1947) gives tables for $3 \leqslant n \leqslant 10$ and $1 \leqslant m_i \leqslant 3$.

the t-distribution is 6·7765. It will be seen that even in this rather extreme case the normal distribution is sufficiently accurate for most purposes. The t-distribution gives in general slightly more accurate values for the tail of the distribution.

It will be shown in a future publication that the test is both consistent and unbiased as the sizes of the samples increase.

My thanks are due to Mr J. W. Whitfield and Mr A. Summerfield for help in the presentation of this paper.

REFERENCES

DAVID, F. N. (1947). *Biometrika*, **34**, 299.
FESTINGER, L. (1946). *Psychometrika*, **11**, 97.
HALDANE, J. B. S. & SMITH, C. A. B. (1948). *Ann. Eugen., Lond.*, **14**, 117.
HOTELLING, H. & PABST, M. R. (1936). *Ann. Math. Statist.* **7**, 29.
IYER, P. V. K. (1952). *J. Indian Soc. Agric. Statist.* **4**, 50.
KENDALL, M. G. (1948). *Rank Correlation Methods*. London: Griffin and Co.
KRUSKAL, W. H. & WALLIS, W. A. (1952). *J. Amer. Statist. Ass.* **47**, 583.
MANN, H. B. & WHITNEY, D. R. (1947). *Ann. Math. Statist.* **18**, 50.
MORAN, P. A. P. (1950). *J. R. Statist. Soc.* B, **12**, 153.
SILVERSTONE, H. (1950). *Biometrika*, **37**, 231.
SILLITO, G. P. (1947). *Biometrika*, **34**, 36.
SUNDRUM, R. M. (1953). *J. R. Statist. Soc.* B, **15**, 246.
VAART, H. R. VAN DER (1950). *Proc. K. Akad. Wet. Amst.* **53**, 494, 507.
WHITNEY, D. R. (1951). *Ann. Math. Statist.* **22**, 274.
WILCOXON, F. (1945). *Biomet. Bull.* **1**, 80.

Notes added in proof. (1) In *Proc. K. Akad. Wet. Amst.* A, **56**, 433 (1953) Terpotra considers the asymptotic normality of S when ties are present in one ranking, under certain conditions which differ from mine, and gives some exact distributions.

(2) *The possibility of ties.* In presenting the test based on S the possibility of ties occurring among the sample values has been excluded. This follows from the assumption that the cumulative distribution functions are continuous and hence that the probability of obtaining a tie is zero. If, in practice, ties should occur *within* a particular sample, then the value of S is not affected; it is only when they occur between the values of two different samples that a difficulty arises. In this case the most stringent action might be to untie the items tied between samples in such a manner that their values become the most unfavourable possible for the alternative hypothesis under consideration. Should the value of S still prove to be significant with this treatment, the null hypothesis may be rejected with some confidence in favour of the alternative. This treatment is of course only feasible if the number of ties between items from different samples is small; should this not be the case it would tend to indicate that the populations from which the samples have been drawn are not continuous and hence invalidate the use, not only of this test but of most others. An adequate treatment of this situation will only be achieved when more is known concerning the small sample distribution of values of S computed from rankings, *both* of which contain ties of any extent and number.

Table 1. *Values of* $\gamma_2 = \kappa_4/\kappa_2^2$ *corresponding to the partitions of ten items into five samples*

m_1	m_2	m_3	m_4	m_5	γ_2
6	1	1	1	1	−0·329
5	2	1	1	1	−0·282
4	3	1	1	1	−0·261
4	2	2	1	1	−0·254
3	3	2	1	1	−0·246
3	2	2	2	1	−0·240
2	2	2	2	2	−0·234

Table 2. *Approximations to the true distribution given by the normal and the t-distribution, for case* $m_1 = 10, m_2 = m_3 = 1, n = 12$

S	Exact probability	Normal approximation		t approximation	
		Probability	Error	Probability	Error
1	0·5000	0·5000	+0·0000	0·5000	0·0000
3	0·4167	0·4154	+0·0013	0·4237	−0·0070
5	0·3409	0·3346	+0·0063	0·3494	−0·0085
7	0·2727	0·2608	+0·0119	0·2795	−0·0068
9	0·2121	0·1964	+0·0157	0·2156	−0·0035
11	0·1591	0·1428	+0·0163	0·1590	+0·0001
13	0·1136	0·1000	+0·0136	0·1110	+0·0026
15	0·0758	0·0674	+0·0084	0·0720	+0·0038
17	0·0455	0·0437	+0·0018	0·0425	+0·0030
19	0·0227	0·0273	−0·0046	0·0219	+0·0008
21	0·0076	0·0163	−0·0087	0·0091	−0·0015

Table 3. *Table giving* $\Pr\{S \geqslant S_0\}$ *for k samples each of size m*

S_0	$k = 3$		S_0	$k = 3$		S_0	$m = 5$ (continued)
	$m = 2$	$m = 4$		$m = 3$	$m = 5$		
0	0·5778	0·5284	1	0·5000	0·5000	51	$0{\cdot}0^3303$
2	0·4222	0·4716	3	0·4155	0·4589	53	$0{\cdot}0^3204$
4	0·2889	0·4156	5	0·3339	0·4182	55	$0{\cdot}0^3134$
6	0·1667	0·3609	7	0·2595	0·3783	57	$0{\cdot}0^3851$
8	0·0889	0·3090	9	0·1940	0·3396	59	$0{\cdot}0^3526$
10	0·0333	0·2602	11	0·1387	0·3025	61	$0{\cdot}0^3313$
12	0·0111	0·2157	13	0·0946	0·2672	63	$0{\cdot}0^3180$
14	—	0·1756	15	0·0613	0·2340	65	$0{\cdot}0^4978$
16	—	0·1404	17	0·0369	0·2032	67	$0{\cdot}0^4502$
18	—	0·1099	19	0·0208	0·1748	69	$0{\cdot}0^4238$
20	—	0·0844	21	0·0107	0·1489	71	$0{\cdot}0^4106$
22	—	0·0632	23	$0{\cdot}0^3476$	0·1256	73	$0{\cdot}0^5396$
24	—	0·0463	25	$0{\cdot}0^3179$	0·1049	75	$0{\cdot}0^5132$
26	—	0·0330	27	$0{\cdot}0^3595$	0·0867		
28	—	0·0229	29	—	0·0708		
30	—	0·0153	31	—	0·0572		
32	—	$0{\cdot}0^2993$	33	—	0·0456		
34	—	$0{\cdot}0^2615$	35	—	0·0359		
36	—	$0{\cdot}0^2367$	37	—	0·0279		
38	—	$0{\cdot}0^3205$	39	—	0·0214		
40	—	$0{\cdot}0^3110$	41	—	0·0161		
42	—	$0{\cdot}0^3519$	43	—	0·0120		
44	—	$0{\cdot}0^3231$	45	—	$0{\cdot}0^3873$		
46	—	$0{\cdot}0^4866$	47	—	$0{\cdot}0^3626$		
48	—	$0{\cdot}0^4289$	49	—	$0{\cdot}0^3440$		

Table 3 (*cont.*)

S_0	$k=4$			$k=5$		$k=6$	S_0
	$m=2$	$m=3$	$m=4$	$m=2$	$m=3$	$m=2$	
0	0·5492	0·5276	0·5183	0·5353	0·5198	0·5271	0
2	0·4508	0·4724	0·4817	0·4647	0·4802	0·4729	2
4	0·3563	0·4177	0·4454	0·3951	0·4408	0·4193	4
6	0·2683	0·3645	0·4094	0·3285	0·4019	0·3670	6
8	0·1929	0·3136	0·3742	0·2667	0·3640	0·3170	8
10	0·1302	0·2659	0·3400	0·2110	0·3273	0·2699	10
12	0·0829	0·2220	0·3069	0·1625	0·2921	0·2265	12
14	0·0484	0·1823	0·2754	0·1213	0·2588	0·1871	14
16	0·0262	0·1472	0·2454	0·0878	0·2274	0·1521	16
18	0·0123	0·1166	0·2172	0·0613	0·1982	0·1215	18
20	$0{\cdot}0^{3}516$	0·0907	0·1910	0·0412	0·1713	0·0953	20
22	$0{\cdot}0^{3}159$	0·0691	0·1666	0·0265	0·1468	0·0734	22
24	$0{\cdot}0^{3}397$	0·0515	0·1443	0·0162	0·1247	0·0553	24
26	—	0·0374	0·1241	$0{\cdot}0^{3}939$	0·1049	0·0408	26
28	—	0·0266	0·1058	$0{\cdot}0^{3}511$	0·0874	0·0294	28
30	—	0·0183	0·0895	$0{\cdot}0^{3}257$	0·0721	0·0207	30
32	—	0·0123	0·0751	$0{\cdot}0^{3}118$	0·0588	0·0142	32
34	—	$0{\cdot}0^{3}797$	0·0624	$0{\cdot}0^{3}476$	0·0475	$0{\cdot}0^{3}943$	34
36	—	$0{\cdot}0^{3}498$	0·0514	$0{\cdot}0^{3}168$	0·0379	$0{\cdot}0^{3}608$	36
38	—	$0{\cdot}0^{3}299$	0·0420	$0{\cdot}0^{4}441$	0·0299	$0{\cdot}0^{3}379$	38
40	—	$0{\cdot}0^{3}171$	0·0339	$0{\cdot}0^{5}882$	0·0234	$0{\cdot}0^{3}227$	40
42	—	$0{\cdot}0^{3}928$	0·0272	—	0·0180	$0{\cdot}0^{3}130$	42
44	—	$0{\cdot}0^{3}471$	0·0215	—	0·0137	$0{\cdot}0^{3}710$	44
46	—	$0{\cdot}0^{3}222$	0·0168	—	0·0102	$0{\cdot}0^{3}366$	46
48	—	$0{\cdot}0^{4}947$	0·0130	—	$0{\cdot}0^{3}755$	$0{\cdot}0^{3}177$	48
50	—	$0{\cdot}0^{4}352$	$0{\cdot}0^{3}998$	—	$0{\cdot}0^{3}549$	$0{\cdot}0^{4}787$	50
52	—	$0{\cdot}0^{4}108$	$0{\cdot}0^{3}754$	—	$0{\cdot}0^{3}392$	$0{\cdot}0^{4}319$	52
54	—	$0{\cdot}0^{5}271$	$0{\cdot}0^{3}562$	—	$0{\cdot}0^{3}275$	$0{\cdot}0^{4}114$	54
56	—	—	$0{\cdot}0^{3}414$	—	$0{\cdot}0^{3}190$	$0{\cdot}0^{3}347$	56
58	—	—	$0{\cdot}0^{3}230$	—	$0{\cdot}0^{3}128$	$0{\cdot}0^{6}802$	58
60	—	—	$0{\cdot}0^{3}214$	—	$0{\cdot}0^{3}848$	$0{\cdot}0^{6}134$	60
62	—	—	$0{\cdot}0^{3}150$	—	$0{\cdot}0^{3}548$	—	62
64	—	—	$0{\cdot}0^{3}104$	—	$0{\cdot}0^{3}345$	—	64
66	—	—	$0{\cdot}0^{3}701$	—	$0{\cdot}0^{3}212$	—	66
68	—	—	$0{\cdot}0^{3}465$	—	$0{\cdot}0^{3}126$	—	68
70	—	—	$0{\cdot}0^{3}301$	—	$0{\cdot}0^{4}724$	—	70
72	—	—	$0{\cdot}0^{3}191$	—	$0{\cdot}0^{4}401$	—	72
74	—	—	$0{\cdot}0^{3}118$	—	$0{\cdot}0^{4}213$	—	74
76	—	—	$0{\cdot}0^{4}705$	—	$0{\cdot}0^{4}108$	—	76
78	—	—	$0{\cdot}0^{4}409$	—	$0{\cdot}0^{5}514$	—	78
80	—	—	$0{\cdot}0^{4}229$	—	$0{\cdot}0^{5}230$	—	80
82	—	—	$0{\cdot}0^{4}122$	—	$0{\cdot}0^{6}945$	—	82
84	—	—	$0{\cdot}0^{5}626$	—	$0{\cdot}0^{6}351$	—	84
86	—	—	$0{\cdot}0^{5}301$	—	$0{\cdot}0^{6}113$	—	86
88	—	—	$0{\cdot}0^{5}136$	—	$0{\cdot}0^{7}297$	—	88
90	—	—	$0{\cdot}0^{5}555$	—	$0{\cdot}0^{8}595$	—	90
92	—	—	$0{\cdot}0^{6}206$	—	—	—	92
94	—	—	$0{\cdot}0^{7}634$	—	—	—	94
96	—	—	$0{\cdot}0^{7}159$	—	—	—	96

TWO-STAGE PROCEDURES FOR ESTIMATING THE DIFFERENCE BETWEEN MEANS†

BY S. G. GHURYE AND HERBERT ROBBINS
University of North Carolina

INTRODUCTION AND SUMMARY

Given two populations P_i ($i = 1, 2$) with unknown means θ_i and variances σ_i^2, we wish to estimate the difference $\theta_1 - \theta_2$. Let $t_i(n)$ be the mean $(X_{i,1} + X_{i,2} + \ldots + X_{i,n})/n$ of a sample from P_i. Then $t_1(n_1) - t_2(n_2)$ is an unbiased estimate of $\theta_1 - \theta_2$, with variance $\sum_i (\sigma_i^2/n_i)$. Assuming the cost of sampling to be a known linear function of the number of observations, the cost of taking n_1 observations from P_1 and n_2 from P_2 is $a_1 n_1 + a_2 n_2 + a_3$. If there is a prescribed upper bound A_0 to the cost of sampling, n_1 and n_2 are subject to the restriction

$$a_1 n_1 + a_2 n_2 \leqslant A = A_0 - a_3. \tag{0·1}$$

The quantity $\sum_i (\sigma_i^2/n_i)$, equal to the variance of $t_1(n_1) - t_2(n_2)$ for integer values of the n_i, is minimized for continuous $n_i > 0$ subject to (0·1) by taking $n_i = n_i^0$, where

$$n_i^0 = (A/a_i)\, a_i^{\frac{1}{2}} \sigma_i / \sum_j a_j^{\frac{1}{2}} \sigma_j; \tag{0·2}$$

the minimum value being equal to

$$V^0(A) = \sum_i (\sigma_i^2/n_i^0) = A^{-1} (\sum_i a_i^{\frac{1}{2}} \sigma_i)^2. \tag{0·3}$$

When the ratio σ_2/σ_1 on which the optimum values (0·2) depend is not known, we can use a two-stage procedure for estimating $\theta_1 - \theta_2$, first taking a sample of $m_1 + m_2$ observations, m_i from P_i, and then using estimates of σ_i obtained from this preliminary sample to distribute the remaining observations between the P_i. We shall investigate the performance of this estimation procedure.

For previous work done on problems of this kind, reference may be made to Putter (1951) and the literature cited in that paper. Putter considers the problem of estimating the mean of a population composed of a known number of normally distributed strata whose relative proportions are known. See also Robbins (1952, p. 528).

In § 1 we assume the P_i to be normal and evaluate the variance of the two-stage estimate. In § 2 we show that as m_1, m_2 and $A \to \infty$ in a certain way, the ratio of this variance to the minimum variance $V^0(A)$ tends to unity, and we also prove the asymptotic result for more general populations.

† Work supported by the U.S. Air Force under Contract AF18(600)–83, monitored by the Office of Scientific Research.

1. NORMAL POPULATIONS

When the P_i are known to be normal, we choose positive integers m_i such that $a_1 m_1 + a_2 m_2 < A$ and take m_i observations from P_i. Let

$$s_i^2(m_i) = \left\{\sum_{j=1}^{m_i} X_{i,j}^2 - m_i t^2(m_i)\right\} \Big/ (m_i - 1) = \text{estimated variance of } P_i, \dagger \tag{1·1}$$

$$u(m_1, m_2) = a_1^{\frac{1}{2}} s_1(m_1) / \sum_i a_i^{\frac{1}{2}} s_i(m_i); \tag{1·2}$$

$$n_1^* = \begin{cases} m_1 & \text{if} \quad Au(m_1, m_2)/a_1 < m_1, \\ (A - a_2 m_2)/a_1 & \text{if} \quad (A - a_2 m_2)/a_1 < Au(m_1, m_2)/a_1, \\ Au(m_1, m_2)/a_1 & \text{if} \quad m_1 \leqslant Au(m_1, m_2)/a_1 \leqslant (A - a_2 m_2)/a_1, \end{cases} \tag{1·3}$$

$$n_2^* = (A - a_1 n_1^*)/a_2, \tag{1·4}$$

and

$$\tilde{n}_i = [n_i^*], \tag{1·5}$$

where $[x]$ is the largest integer in x. Having computed $\tilde{n}_i$ we take $(\tilde{n}_i - m_i)$ more observations $(X_{i,j}, j = m_i + 1, \ldots, \tilde{n}_i)$ from P_i, and estimate $\theta_1 - \theta_2$ by

$$t_1(\tilde{n}_1) - t_2(\tilde{n}_2) = \sum_{j=1}^{\tilde{n}_1} X_{1,j}/\tilde{n}_1 - \sum_{j=1}^{\tilde{n}_2} X_{2,j}/\tilde{n}_2. \tag{1·6}$$

Let

$$\bar{V}(A) = \text{var}\{t_1(\tilde{n}_1) - t_2(\tilde{n}_2)\}. \tag{1·7}$$

Now

$$t_i(\tilde{n}_i) = m_i t_i(m_i)/\tilde{n}_i + (\tilde{n}_i - m_i)\, t_i(\tilde{n}_i - m_i)/\tilde{n}_i, \tag{1·8}$$

where

$$t_i(\tilde{n}_i - m_i) = \sum_{j=m_i+1}^{\tilde{n}_i} X_{i,j}/(\tilde{n}_i - m_i).$$

Since the $\tilde{n}_i$ depend only on the $s_i(m_i)$ for fixed s_i, the random variables $t_1(m_1)$, $t_2(m_2)$, $t_1(\tilde{n}_1 - m_i)$, $t_2(\tilde{n}_2 - m_2)$ are mutually independent, the conditional distributions of $t_i(m_i)$ and $t_i(\tilde{n}_i - m_i)$ being respectively $\mathcal{N}\{\theta_i, \sigma_i^2/m_i\}$ and $\mathcal{N}\{\theta_i, \sigma_i^2/(\tilde{n}_i - m_i)\}$. Hence for fixed s_i the conditional distribution of $t_1(\tilde{n}_i) - t_2(\tilde{n}_2)$ is $\mathcal{N}\{\theta_1 - \theta_2, \sum_i (\sigma_i^2/\tilde{n}_i)\}$. Consequently,

$$\mathrm{E}\{t_1(\tilde{n}_1) - t_2(\tilde{n}_2)\} = \theta_1 - \theta_2 \quad \text{and} \quad \bar{V}(A) = \mathrm{E}\{\sum_i (\sigma_i^2/\tilde{n}_i)\}. \tag{1·9}$$

Let

$$F(u) = \Pr\{u(m_1, m_2) \leqslant u\}. \tag{1·10}$$

Then from (1·5), (1·9) and (1·10) we have

$$\begin{aligned} \bar{V}(A) = {} & \{\sigma_2^2 m_1^{-1} + \sigma_2^2[(A - a_1 m_1)/a_2]^{-1}\} F(a_1 m_1/A) \\ & + \{\sigma_1^2[(A - a_2 m_2)/a_1]^{-1} + \sigma_2^2 m_2^{-1}\}\{1 - F(1 - a_2 m_2/A)\} \\ & + \int_{a_1 m_1/A}^{1 - a_2 m_2/A} \{\sigma_1^2[Au/a_1]^{-1} + \sigma_2^2[A(1-u)/a_2]^{-1}\}\, dF(u). \end{aligned} \tag{1·11}$$

In what follows we shall denote by $V^*(A)$ the expression obtained by dropping the square brackets in (1·11).

Let

$$\left.\begin{aligned} & \rho = \sigma_2/\sigma_1, \quad c = (a_2/a_1)^{\frac{1}{2}}, \quad b_1 = (A - a_1 m_1)/(a_2 m_1), \\ & b_2 = a_1 m_2/(A - a_2 m_2), \quad r_i = (m_i - 1)/2 \quad \text{and} \quad q = r_2/r_1. \end{aligned}\right\} \tag{1·12}$$

† Putter (1951) uses $s_i^2(m_i)\,(m_i - 1)/(m_i - 2)$ instead of $s_i^2(m_i)$ because it minimizes an expression σ_D^2 which is the variance of the estimate obtained by ignoring the fact that the sample sizes prescribed by the two-stage procedure are truncated short of the extreme limits possible under (0·1).

Then from (0·3) we have $$V^0(A) = A^{-1}a_1\sigma_1^2(1+\rho c)^2,$$

and we can reduce the expression for $V^*(A)$ to

$$\begin{aligned} A\{V^*(A) - V^0(A)\} &= a_1\sigma_1^2(\rho - cb_1)^2 b_1^{-1} F(a_1 m_1/A) \\ &\quad + a_1\sigma_1^2(\rho - cb_2)^2 b_2^{-1}\{1 - F(1 - a_2 n_2/A)\} \\ &\quad + \int_{a_1 m_1/A}^{1-a_2 n_2/A} a_1\sigma_1^2\{(1-u)u^{-1} + \rho^2 c^2(1-u)^{-1}u - 2\rho c\}\, dF(u). \end{aligned} \tag{1·13}$$

Finally, making the substitution

$$w = \rho^2 c^2 u^2/\{\rho^2 c^2 u^2 + q(1-u)^2\} = (m_1 - 1)(s_1^2/\sigma_1^2)/\sum_i (m_i - 1)(s_i^2/\sigma_i^2) = \chi^2_{m_1-1}/(\chi^2_{m_1-1} + \chi^2_{m_2-1}), \tag{1·14}$$

so that w has the density function

$$f(w) = \{1/\mathrm{B}(r_1, r_2)\}\, w^{r_1-1}(1-w)^{r_2-1} \quad (0 \leqslant w \leqslant 1), \tag{1·15}$$

we reduce (1·13) to

$$V^*(A)/V^0(A) = 1 + (1+\rho c)^{-2}\{\rho c I_1 + (\rho - cb_1)^2 b_1^{-1} I_2 + (\rho - cb_2)^2 b_2^{-1} I_3\}, \tag{1·16}$$

where

$$\left.\begin{aligned} I_1 &= \int_{\beta_1}^{\beta_2} \{(qw)^{\frac{1}{2}}(1-w)^{-\frac{1}{2}} + (qw)^{-\frac{1}{2}}(1-w)^{\frac{1}{2}} - 2\} f(w)\, dw, \\ I_2 &= \int_0^{\beta_1} f(w)\, dw, \quad I_3 = \int_{\beta_2}^1 f(w)\, dw, \quad \beta_i = \rho^2/\{\rho^2 + c^2 q b_i^2\}. \end{aligned}\right\} \tag{1·17}$$

Hence, V^*/V^0 can be computed by means of tables of the incomplete B-function. We have done this for $a_1 = a_2 = a$, $N = A/a = 30$, 50, $m_1 = m_2 = m = (0·2)N$, $(0·3)N$, $(0·4)N$ and various values of ρ. The results are given in Table 1.

For $a_1 = a_2$, the usual procedure for estimating $\theta_1 - \theta_2$ consists in taking $n_1 = n_2 = \frac{1}{2}N$ and using the estimate $t_1(\frac{1}{2}N) - t_2(\frac{1}{2}N)$. The variance is

$$V' = 2(\sigma_1^2 + \sigma_2^2)/N.$$

For comparing V^* with V', the values of V'/V^0 are given in the last row of the table.

The two-stage procedure seems to effect considerable improvement over the usual one-stage procedure for values of ρ away from 1; and the performance seems to be best for m/N in the neighbourhood of $\sigma_1/(\sigma_1 + \sigma_2)$ if $\sigma_1 < \sigma_2$.

Table 1. *Comparison of V^* with V^0 and V' for normal populations*

	ρ / m/N	1·00	1·25	1·50	1·75	2·00	2·25	2·50	2·75	3·00
V^*/V^0 for $N=30$	0·2	1·064	1·062	1·058	1·054	1·049	1·044	1·039	1·034	1·030
	0·3	1·034	1·032	1·028	1·023	1·018	1·016	1·016	1·018	1·022
	0·4	1·017	1·014	1·012	1·014	1·025	1·039	1·056	1·075	1·094
V^*/V^0 for $N=50$	0·2	1·032	1·031	1·031	1·029	1·027	1·025	1·022	1·019	1·017
	0·3	1·021	1·019	1·017	1·014	1·011	1·010	1·008	1·011	1·016
	0·4	1·013	1·009	1·007	1·010	1·021	1·036	1·055	1·074	1·094
V'/V^0		1·000	1·012	1·040	1·074	1·111	1·148	1·184	1·218	1·250

2. ASYMPTOTIC EFFICIENCY

The idea of substituting $s_i(m_i)$ for σ_i in (0·2) is based on the belief that as $m_i \to \infty$, the ratio $\tilde{n}_1/\tilde{n}_2$ will approach the optimum value n_1^0/n_2^0 and $\tilde{V}/V^0$ will approach unity. We prove that this is the case when the populations are normal, and then we shall prove a similar result which is true also for other populations.

THEOREM 1. Let P_i be normal, and consider $V^*(A)/V^0(A)$ as given by (1·16). Let a_1, a_2 and ρ remain fixed while m_1, m_2 and A become infinite in such a way that

$$\left.\begin{aligned} &0 < h \leqslant m_1/m_2 \leqslant h' < \infty, \quad \text{where } h, h' \text{ are fixed,} \\ \text{and} \quad &m_i/A \to 0 \quad (i = 1, 2). \end{aligned}\right\} \tag{2·1}$$

Then
$$V^*(A)/V^0(A) \to 1. \tag{2·2}$$

Proof. In (1·16),

$$\begin{aligned} I_1 &\leqslant q^{\frac{1}{2}}\mathrm{B}(r_1+\tfrac{1}{2}, r_2-\tfrac{1}{2})/\mathrm{B}(r_1, r_2) + q^{-\frac{1}{2}}\mathrm{B}(r_1-\tfrac{1}{2}, r_2+\tfrac{1}{2})/\mathrm{B}(r_1, r_2) - 2 \\ &= \{q^{\frac{1}{2}}\Gamma(r_1+\tfrac{1}{2})\,\Gamma(r_2-\tfrac{1}{2}) + q^{-\frac{1}{2}}\Gamma(r_1-\tfrac{1}{2})\,\Gamma(r_2+\tfrac{1}{2})\}/\{\Gamma(r_1)\,\Gamma(r_2)\} - 2, \end{aligned}$$

which converges to zero, since

$$x^h\Gamma(x-h)/\Gamma(x) \to 1 \quad \text{as} \quad x \to \infty. \tag{2·3}$$

$$\begin{aligned} I_2 &\leqslant \Pr\{w < \rho^2/(\rho^2 + c^2qb_1^2)\} \leqslant \Pr\{w^{-1} > c^2qb_1^2/\rho^2\} \\ &\leqslant \mathrm{E}\{w^{-1}\}\rho^2/(c^2qb_1^2), \end{aligned}$$

and

$$\begin{aligned} I_3 &\leqslant \Pr\{w > \rho^2/(\rho^2 + c^2qb_2^2)\} \leqslant \Pr\{(1-w)^{-1} > \rho^2/(c^2qb_2^2)\} \\ &\leqslant \mathrm{E}\{(1-w)^{-1}\}c^2qb_2^2/\rho^2. \end{aligned}$$

Now
$$\mathrm{E}\{w^{-1}\} = \Gamma(r_1-1)\,\Gamma(r_1+r_2)/\{\Gamma(r_1)\,\Gamma(r_1+r_2-1)\},$$

and
$$\mathrm{E}\{(1-w)^{-1}\} = \Gamma(r_2-1)\,\Gamma(r_1+r_2)/\{\Gamma(r_2)\,\Gamma(r_1+r_2-1)\},$$

both of which remain bounded on account of (2·1) and (2·4). Since $b_1 \to \infty$ and $b_2 \to 0$, we have

$$(\rho - cb_1)^2 b_1^{-1} I_2 \to 0 \quad \text{and} \quad (\rho - cb_2)^2 b_2^{-1} I_3 \to 0.$$

Hence (2·2) is proved.

From the expression (1·11) for $\tilde{V}(A)$, it follows that

$$\tilde{V}(A)/V^0(A) \to 1.$$

Next, we remove the restriction that the P_i be normal, assuming only that they have finite variances σ_i^2 and that we know functions f_i such that the statistics

$$s_i(n) = f_i(X_{i,1}, \ldots, X_{i,n};\, n) \quad (i = 1, 2) \tag{2·4}$$

satisfy conditions (I)–(III) below.

Let
$$F_i(s;\, n) = \Pr\{s_i(n) \leqslant s\} \quad \text{and} \quad C_i(\epsilon) = [\sigma_i - \epsilon, \sigma_i + \epsilon].$$

We assume:

(I) There exists an $\alpha > 1$ such that for every fixed $\epsilon > 0$,

$$n^\alpha \Pr\{s_i(n) \notin C_i(\epsilon)\} \text{ is bounded for all } n > 0,\ i = 1, 2;$$

(II) There exists an $\epsilon > 0$ and $< \min(\sigma_1, \sigma_2)$ such that $n\left\{\int_{C_i(\epsilon)} s^k dF_i(s;\, n) - \sigma_i^k\right\}$ is bounded for all $n > 0$ and $k = 2, -2$;

(III) Either $t_i(n)$ and $s_i(n)$ from the same sample are a pair of mutually independent random variables, or P_i has a finite fourth moment.

We shall follow the two-stage procedure given by (1·2)–(1·6) with $s_i(m_i)$ as given by (2·4) instead of (1·1). Then we have

THEOREM 2. Let a_1, a_2 and ρ remain fixed while m_1, m_2 and A become infinite in such a way that

$$(2\cdot1) \text{ holds, and } A/m^{k(a+1)} \text{ is bounded.} \tag{2·5}$$

Then
$$E\{t_1(\tilde{n}_1)-t_2(\tilde{n}_2)\}\to\theta_1-\theta_2, \tag{2·6}$$

$$\tilde{V}(A)/V^0(A)\to 1. \tag{2·7}$$

Proof. We shall first prove the following statements:

$$\text{Assumption (II) also holds for } k=1 \text{ and } -1; \tag{2·8}$$

$$Am_i\{E(\tilde{n}_i^{-k})-(n_i^0)^{-k}\} \text{ is bounded for } k=1,2; \tag{2·9}$$

$$A^2m_i^3E(\tilde{n}_i^{-4})\to 0. \tag{2·10}$$

To prove (2·8) for $k=-1$, we note that

$$\int_{C_i(\varepsilon)} s^{-1}dF_i(s;n)\leqslant\left\{\int_{C_i(\varepsilon)} s^{-2}dF_i(s;n)\right\}^{\frac{1}{2}}=\sigma_i^{-1}\{1+O(n^{-1})\};$$

and since
$$s_i^{-1}=\sigma_i^{-1}\{1+(s_i^2\sigma_i^{-2}-1)\}^{-\frac{1}{2}}\geqslant\sigma_i^{-1}\{1-(\tfrac{1}{2})(s_i^2\sigma_i^{-2}-1)\}$$

for $s_i\in C_i(\varepsilon)$, we have

$$\int_{C_i(\varepsilon)} s^{-1}dF_i(s;n)\geqslant\sigma_i^{-1}\left\{(\tfrac{3}{2})\int_{C_i(\varepsilon)}dF_i(s;n)-(\tfrac{1}{2})\int_{C_1(\varepsilon)}s^2\sigma_i^{-2}dF_i(s;n)\right\}$$

$$=\sigma_i^{-1}\{1+O(n^{-1})\}, \quad \text{by (I) and (II).}$$

Hence (2·8) is proved for $k=-1$, and in the same way for $k=1$. We need only prove (2·9) and (2·10) for $i=1$, since the proof for $i=2$ is similar. Moreover, from the definitions of $\tilde{n}_i$ and n_i^*, it is evident that (2·9) and (2·10) hold if and only if they hold with $\tilde{n}_i$ replaced by n_i^*.

Let $m=(m_1,m_2)$ and set
$$v_m=s_2(m_2)/s_1(m_1), \tag{2·11}$$

and
$$v_m^*=\begin{cases}N_m & \text{if } v_m>N_m=(A-a_1m_1)/(ca_1m_1),\\ \delta_m & \text{if } v_m<\delta_m=a_2m_2/(cA-ca_2m_2),\\ v_m & \text{if } \delta_m\leqslant v_m\leqslant N_m.\end{cases} \tag{2·12}$$

Then
$$\delta_m=O(m_iA^{-1})\to 0 \quad \text{and} \quad N_m=O(m_i^{-1}A)\to\infty, \tag{2·13}$$

and
$$(n_i^*)^{-1}=a_1(1+cv_m^*)/A. \tag{2·14}$$

Consequently, to prove (2·9), we need only show that

$$m_1\{E(v_m^*)^k-\rho^k\} \text{ is bounded for } k=1,2. \tag{2·15}$$

Let us choose an ε to satisfy (II); we can, by (2·13), choose m_1, m_2, A large enough so that

$$\delta_m\leqslant(\sigma_2-\varepsilon)/(\sigma_1+\varepsilon) \quad \text{and} \quad N_m\geqslant(\sigma_2+\varepsilon)/(\sigma_1-\varepsilon),$$

and hence such that
$$\{s_i(m_i)\in C_i(\varepsilon) \quad (i=1,2)\}\Rightarrow\delta_m\leqslant v_m\leqslant N_m.$$

Under these circumstances, we have from (2·12)

$$0 \leqslant E(v_m^*)^k - \int_{C_1(\epsilon)} \int_{C_2(\epsilon)} s_1^{-k} s_2^k \, dF_2(s_2; m_2) \, dF_1(s_1; m_1) \leqslant N_m^k p(m_1, m_2), \tag{2·16}$$

where $$p(m_1, m_2) = 1 - \prod_{i=1,2} \Pr\{s_i(m_i) \in C_i(\epsilon)\} = O(m^{-\alpha}).$$

By (II), the second term in the middle in (2·16) is $\rho^k + O(m^{-1})$, and by (2·13) the last term is $O(A^k m^{-k} m^{-\alpha}) = O(m^{-1})$ for $k = 1, 2$. Thus, we have (2·15) and hence (2·9). We can prove (2·10) similarly.

Now let
$$\left.\begin{aligned} T_1 &= m_1 t_1(m_1)/\tilde{n}_1 - m_2 t_2(m_2)/\tilde{n}_2, \\ T_2 &= (\tilde{n}_1 - m_1)\, t_1(\tilde{n}_1 - m_1)/\tilde{n}_1 - (\tilde{n}_2 - m_2)\, t_2(\tilde{n}_2 - m_2)/\tilde{n}_2, \end{aligned}\right\} \tag{2·17}$$

where $$(\tilde{n}_i - m_i)\, t_i(\tilde{n}_i - m_i) = \sum_{j=m_i+1}^{\tilde{n}_i} X_{i,j}.$$

Then $$E\{t_1(\tilde{n}_1) - t_2(\tilde{n}_2)\} = E\{T_1\} + E_{\tilde{n}_1} E\{T_2 \mid \tilde{n}_1\}. \tag{2·18}$$

Since $\tilde{n}_i$ depends only on $\{X_{i,j}, j = 1, 2, \ldots, m_i\}$ and $t_i(\tilde{n}_i - m_i)$ for fixed $\tilde{n}_1$, only on

$$\{X_{i,j}, j = m_i + 1, \ldots, \tilde{n}_i\},$$

we have $$\Pr\{t_i(\tilde{n}_i - m_i) \leqslant x_i, \; i = 1, 2 \mid \tilde{n}_i = n_i\} = \prod_{i=1,2} \Pr\{t_i(n_i - m_i) \leqslant x_i\}. \tag{2·19}$$

Therefore $$E\{T_2 \mid \tilde{n}_1\} = (\tilde{n}_1 - m_1)\, \theta_1/\tilde{n}_1 - (\tilde{n}_2 - m_2)\, \theta_2/\tilde{n}_2,$$

so that
$$\begin{aligned} E_{\tilde{n}_1} E\{T_2 \mid \tilde{n}_1\} &= \theta_1 - \theta_2 - m_1 \theta_1 E(1/\tilde{n}_1) + m_2 \theta_2 E(1/\tilde{n}_2) \\ &\to \theta_1 - \theta_2 \quad \text{by (2·5) and (2·9).} \end{aligned}$$

Moreover,
$$\begin{aligned} m_i E\{t_i(m_i)/\tilde{n}_i\} &\leqslant m_i \{E t_i^2(m_i)\, E(1/\tilde{n}_i^2)\}^{\frac{1}{2}} \\ &= A^{-1} m_i O(1) \{A^2 E(1/\tilde{n}_i^2)\}^{\frac{1}{2}} \\ &\to 0 \quad \text{by (2·5) and (2·9).} \end{aligned}$$

Therefore, $ET_1 \to 0$, and hence we have (2·6).

Finally, $$A\, \tilde{V}(A) = A \operatorname{var}(T_1 + T_2) = A \operatorname{var} T_1 + A \operatorname{var} T_2 + 2A \operatorname{cov}(T_1, T_2). \tag{2·20}$$

$$\begin{aligned} A \operatorname{var} T_1 &= A \operatorname{var}\{m_1 t_1(m_1)/\tilde{n}_1 - m_2 t_2(m_2)/\tilde{n}_2\} \\ &\leqslant 2A m_1^2 \operatorname{var}\{t_1(m_1)/\tilde{n}_1\} + 2A m_2^2 \operatorname{var}\{t_2(m_2)/\tilde{n}_2\} \\ &\leqslant 2 \sum_i A m_i^2 \operatorname{var}\{[t_i(m_i) - \theta_i]/\tilde{n}_i + \theta_i/\tilde{n}_i\} \\ &\leqslant 4 \sum_i A m_i^2 E\{[t_i(m_i) - \theta_i]^2/\tilde{n}_i^2\} + 4 \sum_i A m_i^2 \theta_i^2 \operatorname{var}(1/\tilde{n}_i). \end{aligned} \tag{2·21}$$

From (2·5) and (2·9) we know that $A m_i^2 \operatorname{var}(1/\tilde{n}_i) \to 0$. As for the other term on the right-hand side in (2·21), if $t_i(m_i)$ and $s_i(m_i)$ are independent, we have

$$\begin{aligned} A m_i^2 E\{[t_i(m_i) - \theta_i]^2/\tilde{n}_i^2\} &= A m_i^2 E\{[t_i(m_i) - \theta_i]^2\} E(1/\tilde{n}_i^2) \\ &= A m_i \dot{\sigma}_i^2 E(1/\tilde{n}_i^2) \\ &\to 0 \quad \text{by (2·9).} \end{aligned}$$

If $t_i(m_i)$ and $s_i(m_i)$ are not independent, we still have

$$\begin{aligned} Am_i^2 E\{[t_i(m_i)-\theta_i]^2/\tilde{n}_i^2\} &\leqslant Am_i^2\{E[t_i(m_i)-\theta_i]^4\, E(1/\tilde{n}_i^4)\}^{\frac{1}{2}} \\ &= \{m_i^2 E[t_i(m_i)-\theta_i]^4\, A^2 m_i^2 E(1/\tilde{n}_i^4)\}^{\frac{1}{2}} \\ &\to 0 \quad \text{by (III) and (2·10).} \end{aligned}$$

Hence, $A \operatorname{var}(T_1) \to 0$.

We shall see below that $A \operatorname{var}(T_2)$ is bounded, so that

$$A \operatorname{cov}(T_1, T_2) \leqslant \{A \operatorname{var}(T_1)\, A \operatorname{var}(T_2)\}^{\frac{1}{2}} \to 0.$$

Therefore

$$\begin{aligned} \lim A\bar{V}(A) &= \lim A \operatorname{var} T_2 = \lim AE_{\tilde{n}_1} \operatorname{var}(T_2 \mid \tilde{n}_1) + \lim A \operatorname{var} E(T_2 \mid \tilde{n}_1) \\ &= \lim AE_{\tilde{n}_1} \operatorname{var}\{(\tilde{n}_1 - m_1)\, t_1(\tilde{n}_1 - m_1)/\tilde{n}_1 \\ &\quad - (\tilde{n}_2 - m_2)\, t_2(\tilde{n}_2 - m_2)/\tilde{n}_2\} \\ &\quad + \lim A \operatorname{var}\{(\tilde{n}_1 - m_1)\, \theta_1/\tilde{n}_1 - (\tilde{n}_2 - m_2)\, \theta_2/\tilde{n}_2\} \\ &= \lim AE_{\tilde{n}_1}\{\textstyle\sum_i (\tilde{n}_i - m_i)\, \sigma_i^2/\tilde{n}_i^2\} + \lim A \operatorname{var}\{m_2\theta_2/\tilde{n}_2 - m_1\theta_1/\tilde{n}_1\} \\ &= \lim \textstyle\sum_i \sigma_i^2 AE(1/\tilde{n}_i) - \lim \sum_i Am_i\sigma_i^2 E(1/\tilde{n}_i^2) + \lim A \operatorname{var}\{m_2\theta_2/\tilde{n}_2 - m_1\theta_1/\tilde{n}_1\}. \end{aligned} \tag{2·22}$$

The last two terms are zero on account of (2·5) and (2·9); and from (2·9) we see that the first term on the right-hand side of (2·22) is the required limit in (2·7).

If we use sample sizes $n_i' = (A/a_i)\, a_i^{\frac{1}{2}} / \sum_j a_j^{\frac{1}{2}}$, which by (0·2) is what we would be led to do if we thought that $\sigma_1 = \sigma_2$, the variance of the estimate of $\theta_1 - \theta_2$ would be

$$V' = V^0\{1 + c(1-\rho)^2/(1+\rho c)^2\} > V^0 \quad \text{for} \quad \rho > 1.$$

Hence, asymptotically, the two-stage procedure is more efficient than this one-stage procedure if $\rho > 1$.

EXAMPLES. (1) If the P_i are normal, the conditions of Theorem 2 are satisfied for every $\alpha > 0$, and hence in (2·5) A may increase as any power of m_i. We have seen in Theorem 1 that it is actually not necessary to restrict A to be of the order of a power of m_i.

(2) If the P_i are Poisson, $\sigma_i^2 = \theta_i$ and $s_i^2(n) = t_i(n)$. From the fact that $nt_i(n)$ has a Poisson distribution, it can be seen that the conditions of Theorem 2 are satisfied for every $\alpha > 0$.

(3) If the P_i are binomial, with $\Pr\{X_i = 1\} = \theta_i$ and $\Pr\{X_i = 0\} = 1 - \theta_i$, we have $\sigma_i^2 = \theta_i(1-\theta_i)$ and $s_i^2(n) = t_i(n)\{1 - t_i(n)\}$. Using the fact that $nt_i(n)$ is a binomial variate, we can show that the conditions of the theorem are satisfied for every $\alpha > 0$.

(4) If we do not know the forms of P_i, we would use the estimate $s_i(m_i)$ of σ_i given by (1·1). If we know that P_i has a sufficient number, say 8, of moments finite, we can show that Theorem 2 is true for the procedure given by (1·1)–(1·6).

REFERENCES

PUTTER, J. (1951). Sur une methode de double échantillonnage, etc. *Rev. Inst. Int. Statist.* part 3, 1–8.

ROBBINS, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **58**, 527–35.

A BIVARIATE GENERALIZATION OF STUDENT'S t-DISTRIBUTION, WITH TABLES FOR CERTAIN SPECIAL CASES†

BY CHARLES W. DUNNETT‡ AND MILTON SOBEL

Cornell University

A multivariate generalization of Student's t-distribution is considered. The bivariate case is treated in detail; exact and asymptotic expressions for the probability integral and an asymptotic expression for certain percentage points are obtained. The main results for the bivariate case are given as equations (10), (11), (23) and (30) below; these equations are used to construct tables for certain special cases.

1. INTRODUCTION

Consider the joint distribution of p variates $t_i = z_i/s$ $(i = 1, 2, \ldots, p)$, the z_i having a non-singular multivariate normal distribution with means 0, common unknown variance σ^2 and known correlation matrix $\{\rho_{ij}\}$ and ns^2/σ^2 having a χ^2_n-distribution, independent of the z_i, with n degrees of freedom.

The joint density of $z_1, \ldots, z_p, s$ is clearly

$$\frac{2A^{\frac{1}{2}}}{(n\pi)^{\frac{1}{2}p}\,\Gamma(\frac{1}{2}n)}\,(n/2\sigma^2)^{\frac{1}{2}(n+p)}\,s^{n-1}\left\{\exp -\frac{1}{2}\left(\sum_{i,j=1}^{p} a_{ij}z_iz_j + ns^2\right)\Big/\sigma^2\right\}dz_1 \ldots dz_p\,ds, \tag{1}$$

where A is the determinant of the positive definite matrix $\{a_{ij}\} = \{\rho_{ij}\}^{-1}$. Letting $z_i = t_i s$ $(i = 1, 2, \ldots, p)$ and integrating out s, we obtain the expression

$$\frac{A^{\frac{1}{2}}\Gamma[\frac{1}{2}(n+p)]}{(n\pi)^{\frac{1}{2}p}\,\Gamma(\frac{1}{2}n)}\left[1 + \frac{1}{n}\sum_{i,j} a_{ij}t_it_j\right]^{-\frac{1}{2}(n+p)} dt_1 \ldots dt_p \tag{2}$$

for the joint density of $t_1, t_2, \ldots, t_p$. As $n \to \infty$, (2) approaches the standardized p-variate normal density with correlation matrix $\{\rho_{ij}\}$. For $p = 1$, it reduces to Student's t-distribution with n degrees of freedom. Thus it may be considered as a multivariate generalization of Student's t-distribution.

The distribution defined by (2) arose in connexion with certain multiple decision problems concerned with the ranking of $p+1$ populations according to their population means when the population distributions are assumed to be normal with a common unknown variance. These applications are discussed in a separate paper (Bechhofer, Dunnett & Sobel, 1954).

The distribution (2) with $\rho_{ij} = \frac{1}{2}(i \neq j)$ also can be used in connexion with a multiple decision problem considered by Paulson (1952). The symbol λ_n defined on the top of p. 244 of his paper is, in our terminology, the equi-coordinate percentage point (defined for $p = 2$ in § 2 of this paper); for $k = 3$ $(p = 2)$ these percentage points are given as $t_n(P; \frac{1}{2})$ in Table 3.

The major purpose of the present paper is to derive expressions for the probability integral and the equi-coordinate percentage points for the bivariate case of (2). Certain useful tables computed from these expressions are presented. The problem of computing equi-coordinate percentage points for the multivariate case of (2), using Table 1, will be discussed in a further Note which it is hoped to publish in the next issue of this Journal.

† This research was sponsored by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command.

‡ Now at Lederle Laboratories, Pearl River, New York.

2. Exact expression for the bivariate probability integral

For the bivariate case ($p = 2$), it will be convenient to replace t_1, t_2 and ρ_{12} by u, v and ρ, respectively. Then (2) reduces to

$$g_n(u, v; \rho) = \frac{1}{2\pi\sqrt{(1-\rho^2)}}\left[1+\frac{u^2-2\rho uv+v^2}{n(1-\rho^2)}\right]^{-\frac{1}{2}(n+2)}. \tag{3}$$

We wish to evaluate the probability integral

$$P_n(h, k; \rho) = \int_{-\infty}^{k}\int_{-\infty}^{h} g_n(u, v; \rho)\, du\, dv. \tag{4}$$

The integral of (3) over any region of the type $h_1 < u < h_2$, $k_1 < v < k_2$, finite or infinite, can be expressed in terms of (4) and the probability integral of the univariate Student t-distribution.

As a first step in the evaluation of (4) we make the transformation

$$\left.\begin{aligned} r\cos\theta &= \frac{u-\rho v}{\sqrt{(1-\rho^2)}}, \\ r\sin\theta &= v, \end{aligned}\right\} \tag{5}$$

the Jacobian of which is $r\sqrt{(1-\rho^2)}$. We obtain

$$\begin{aligned} P_n(h, k; \rho) &= \tfrac{1}{4}(1+\operatorname{sgn} h)(1+\operatorname{sgn} k) \\ &\quad - \int_{C(h,k;\rho)}^{\pi}\int_{k\csc\theta}^{\infty}\phi_n(r)\, dr\, d\theta - \int_{C(k,h;\rho)}^{\pi}\int_{h\csc\theta}^{\infty}\phi_n(r)\, dr\, d\theta, \end{aligned} \tag{6}$$

where $$\phi_n(r) = \frac{r}{2\pi}\left(1+\frac{r^2}{n}\right)^{-\frac{1}{2}(n+2)}, \quad C(h, k; \rho) = \arctan\frac{k\sqrt{(1-\rho^2)}}{h-\rho k},$$

and $$\operatorname{sgn} m = \begin{cases} +1 & \text{if} \quad m \geqq 0, \\ -1 & \text{if} \quad m < 0. \end{cases}$$

In (6) and throughout the remainder of the paper we assume, unless stated otherwise, that for any real A, B we have

$$0 \leqq \arctan\frac{A}{B} \leqq 2\pi,$$

and the angle is to be interpreted as lying in the interval $(0, \frac{1}{2}\pi)$, $(\frac{1}{2}\pi, \pi)$, $(\pi, \frac{3}{2}\pi)$, or $(\frac{3}{2}\pi, 2\pi)$ according as the signs of (A, B) are $(+, +)$, $(+, -)$, $(-, -)$, or $(-, +)$. For $h = k = 0$, the lower limits on θ are indeterminate, but if we set $h/k = \xi$ the limiting value $P_n(0, 0; \rho)$, given in (15) below, is independent of ξ. If $\rho = \pm 1$ and $h = \rho k$, we define $\arctan\dfrac{k\sqrt{(1-\rho^2)}}{h-\rho k}$ to be $\frac{1}{2}\pi$ if k is positive and $\frac{3}{2}\pi$ if k is negative and similarly for $\arctan\dfrac{h\sqrt{(1-\rho^2)}}{k-\rho h}$.

The function $\phi_n(r)$ is immediately integrable with respect to r, and if we define

$$Q_{\frac{1}{2}n}(m, h, k) = \frac{1}{2\pi}\int_{C(h,k;\rho)}^{\pi}\left(1+\frac{k^2\csc^2\theta}{m}\right)^{-\frac{1}{2}n} d\theta, \tag{7}$$

then the result of the integration can be expressed as

$$P_n(h, k; \rho) = \tfrac{1}{4}(1+\operatorname{sgn} h)(1+\operatorname{sgn} k) - Q_{\frac{1}{2}n}(n, h, k) - Q_{\frac{1}{2}n}(n, k, h). \tag{8}$$

The recursion formula

$$Q_{\frac{1}{2}n}(m,h,k) = Q_{\frac{1}{2}n-1}(m,h,k) - \frac{k}{4\sqrt{(m\pi)}}\frac{1}{(1+k^2/m)^{\frac{1}{2}(n-1)}}\frac{\Gamma[\frac{1}{2}(n-1)]}{\Gamma(\frac{1}{2}n)}\{1+\operatorname{sgn}(h-\rho k)\, I_{x(m,h,k)}[\tfrac{1}{2},\tfrac{1}{2}(n-1)]\} \quad (9)$$

is useful in reducing (8). Here

$$x(m,h,k) = \frac{(h-\rho k)^2}{(h-\rho k)^2+(1-\rho^2)(m+k^2)}$$

and

$$I_{x(m,h,k)}(p,q) = \int_0^{x(m,h,k)} \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)} y^{p-1}(1-y)^{q-1}\,dy$$

is the incomplete beta function tabulated by Pearson (1931). If we make repeated applications of (9) in (8), and compute $Q_0(n,h,k)+Q_0(n,k,h)$ and $Q_{\frac{1}{2}}(n,h,k)+Q_{\frac{1}{2}}(n,k,h)$, we obtain for even n

$$\begin{aligned} P_n(h,k;\rho) = {} & \frac{1}{2\pi}\arctan\frac{\sqrt{(1-\rho^2)}}{-\rho} \\ & + \frac{k}{4\sqrt{(n\pi)}}\sum_{j=1}^{\frac{1}{2}n}\frac{\Gamma(j-\frac{1}{2})}{\Gamma(j)}\frac{1}{(1+k^2/n)^{j-\frac{1}{2}}}[1+\operatorname{sgn}(h-\rho k)\, I_{x(n,h,k)}(\tfrac{1}{2},j-\tfrac{1}{2})] \\ & + \frac{h}{4\sqrt{(n\pi)}}\sum_{j=1}^{\frac{1}{2}n}\frac{\Gamma(j-\frac{1}{2})}{\Gamma(j)}\frac{1}{(1+h^2/n)^{j-\frac{1}{2}}}[1+\operatorname{sgn}(k-\rho h)\, I_{x(n,k,h)}(\tfrac{1}{2},j-\tfrac{1}{2})], \end{aligned} \quad (10)$$

and for odd n

$$\begin{aligned} P_n(h,k;\rho) = {} & \frac{1}{2\pi}\arctan\left\{\sqrt{n}\left[\frac{-(h+k)(hk+\rho n)-(hk-n)\sqrt{(h^2-2\rho hk+k^2+n[1-\rho^2])}}{(hk-n)(hk+\rho n)-n(h+k)\sqrt{(h^2-2\rho hk+k^2+n[1-\rho^2])}}\right]\right\} \\ & + \frac{k}{4\sqrt{(n\pi)}}\sum_{j=1}^{\frac{1}{2}(n-1)}\frac{\Gamma(j)}{\Gamma(j+\frac{1}{2})}\frac{1}{(1+k^2/n)^j}[1+\operatorname{sgn}(h-\rho k)\, I_{x(n,h,k)}(\tfrac{1}{2},j)] \\ & + \frac{h}{4\sqrt{(n\pi)}}\sum_{j=1}^{\frac{1}{2}(n-1)}\frac{\Gamma(j)}{\Gamma(j+\frac{1}{2})}\frac{1}{(1+h^2/n)^j}[1+\operatorname{sgn}(k-\rho h)\, I_{x(n,k,h)}(\tfrac{1}{2},j)]. \end{aligned} \quad (11)$$

The expression in the first term of (11) can be rationalized and simplified, but the exact rule of determining the quadrant then becomes involved.

For computational purposes, it is not necessary to use Pearson's tables in evaluating the incomplete beta functions. Using the reduction formula

$$I_x(p,q) = I_x(p-1,q) - \frac{\Gamma(p+q-1)}{\Gamma(p)\,\Gamma(q)}x^{p-1}(1-x)^q, \quad (12)$$

which is easily derived, we obtain

$$I_x(\tfrac{1}{2},j-\tfrac{1}{2}) = \frac{2}{\pi}\arctan\sqrt{\left(\frac{x}{1-x}\right)} + \frac{2}{\pi}\sqrt{\{x(1-x)\}}\sum_{i=0}^{j-2}\frac{4^i(i!)^2}{(2i+1)!}(1-x)^i \quad (13)$$

and

$$I_x(\tfrac{1}{2},j) = \sqrt{x}\sum_{i=0}^{j-1}\frac{(2i)!}{4^i(i!)^2}(1-x)^i. \quad (14)$$

It is of interest to note that the result for $h=k=0$ in both (10) and (11) is

$$P_n(0,0;\rho) = \arctan\frac{\sqrt{(1-\rho^2)}}{-\rho}, \quad (15)$$

which is independent of n and is therefore identical with the corresponding result for the bivariate normal integral.

3. Asymptotic expression for the bivariate probability integral

Since the number of terms in (10) and (11) above increases with n, the usefulness of these expressions is limited to small values of n. An asymptotic series in powers of n^{-1} will now be derived, the first few terms of which will yield a good approximation to the probability integral even for moderately small values of n. The method of derivation is essentially the same as that used by Fisher (1925) to approximate the probability integral of the univariate Student t-distribution.

Consider the probability density function $g_n(u, v; \rho)$ defined by (3), and the limiting normal form

$$g(u, v; \rho) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left\{-\frac{u^2 - 2\rho uv + v^2}{2(1-\rho^2)}\right\}, \tag{16}$$

obtained when $n \to \infty$. We will express the difference $g_n(u, v; \rho) - g(u, v; \rho)$ as a power series in n^{-1} and then integrate this series term by term over the desired region of integration. The desired integral of (3) then is equal to the integral of (16), which has been tabulated by Pearson (1931), plus a series of correction terms.

It is convenient to let

$$r^2 = \frac{u^2 - 2\rho uv + v^2}{1-\rho^2}, \tag{17}$$

which is in accordance with the notation introduced in (5). A straightforward logarithmic expansion in powers of n^{-1} gives

$$-\left(\frac{n+2}{2}\right)\log\left(1+\frac{r^2}{n}\right) = -\frac{r^2}{2} + \left(\frac{r^4}{4} - r^2\right)\frac{1}{n} + \left(\frac{r^4}{2} - \frac{r^6}{6}\right)\frac{1}{n^2} + \left(\frac{r^8}{8} - \frac{r^6}{3}\right)\frac{1}{n^3} + \left(\frac{r^8}{4} - \frac{r^{10}}{10}\right)\frac{1}{n^4} + \ldots. \tag{18}$$

Hence, by (3) and (16), we can write

$$\begin{aligned}\frac{g_n(u, v; \rho)}{g(u, v; \rho)} &= \exp\left\{\left(\frac{r^4}{4} - r^2\right)\frac{1}{n} + \left(\frac{r^4}{2} - \frac{r^6}{6}\right)\frac{1}{n^2} + \left(\frac{r^8}{8} - \frac{r^6}{3}\right)\frac{1}{n^3} + \left(\frac{r^8}{4} - \frac{r^{10}}{10}\right)\frac{1}{n^4} + \ldots\right\} \\ &= 1 + \left(\frac{r^4}{4} - r^2\right)\frac{1}{n} + \left(\frac{r^8}{32} - \frac{5r^6}{12} + r^4\right)\frac{1}{n^2} + \left(\frac{r^{12}}{384} - \frac{7r^{10}}{96} + \frac{13r^8}{24} - r^6\right)\frac{1}{n^3} \\ &\qquad + \left(\frac{r^{16}}{6144} - \frac{r^{14}}{128} + \frac{17r^{12}}{144} - \frac{77r^{10}}{120} + r^8\right)\frac{1}{n^4} + \ldots \\ &= 1 + D(r), \quad \text{say.}\end{aligned} \tag{19}$$

Then the desired probability integral satisfies the relationship

$$\int_{-\infty}^{k}\int_{-\infty}^{h} [g_n(u, v; \rho) - g(u, v; \rho)]\, du\, dv = \int_{-\infty}^{k}\int_{-\infty}^{h} D(r)\, g(u, v; \rho)\, du\, dv. \tag{20}$$

Applying the transformation (5), and integrating with respect to r, the right-hand side of (20) can be written, after some algebraic simplification, as

$$-\int_{C(h,k;\rho)}^{\pi} F(k, k\cot\theta)\, k\csc^2\theta\, d\theta - \int_{C(k,h;\rho)}^{\pi} F(h, h\cot\theta)\, h\csc^2\theta\, d\theta, \tag{21}$$

where

$$F(k,x)=\frac{k}{4\pi}e^{-\frac{1}{2}k(x^2+k^2)}\left\{\frac{1}{2n}[x^2+k^2]+\frac{1}{48n^2}[3x^6+(9k^2-16)x^4\right.$$
$$+(9k^4-32k^2)x^2+(3k^6-16k^4)]+\frac{1}{192n^3}[x^{10}+(5k^2-16)x^8+(10k^4-64k^2+48)x^6$$
$$+(10k^6-96k^4+144k^2)x^4+(5k^8-64k^6+144k^4)x^2+(k^{10}-16k^8+48k^6)]$$
$$+\frac{1}{46080n^4}[15x^{14}+(105k^2-480)x^{12}+(315k^4-2880k^2+4160)x^{10}$$
$$+(525k^6-7200k^4+20800k^2-9216)x^8+(525k^8-9600k^6+41600k^4-36864k^2)x^6$$
$$+(315k^{10}-7200k^8+41600k^6-55296k^4)x^4+(105k^{12}-2880k^{10}$$
$$\left.+20800k^8-36864k^6)x^2+(15k^{14}-480k^{12}+4160k^{10}-9216k^8)]+\ldots\right\}. \tag{22}$$

If we make the transformation $w=k\cot\theta$ in the first integral of (21), and $w=h\cot\theta$ in the second integral, and integrate with respect to w in each of them, then our final result, after further algebraic simplification, becomes

$$\int_{-\infty}^{k}\int_{-\infty}^{h}g_n(u,v;\rho)\,du\,dv=\int_{-\infty}^{k}\int_{-\infty}^{h}g(u,v;\rho)\,du\,dv+F^*(a,k)+F^*(b,h), \tag{23}$$

where

$$F^*(a,k)=-\tfrac{1}{2}kH(k)\,G(a)\left\{\frac{1}{2n}(k^2+1)+\frac{1}{48n^2}(3k^6-7k^4-5k^2-3)\right.$$
$$+\frac{1}{192n^3}(k^{10}-11k^8+14k^6+6k^4-3k^2-15)$$
$$\left.+\frac{1}{46080n^4}(15k^{14}-375k^{12}+2225k^{10}-2141k^8-939k^6-213k^4+915k^2+945)+\ldots\right\}$$
$$+\tfrac{1}{2}akH(a)\,H(k)\left\{\frac{1}{2n}+\frac{1}{48n^2}[9k^4+k^2(9a^2-5)+(3a^4-a^2-3)]\right.$$
$$+\frac{1}{192n^3}[5k^8+k^6(10a^2-34)+k^4(10a^4-46a^2+6)+k^2(5a^6-29a^4-a^2-3)$$
$$+(a^8-7a^6-a^4-5a^2-15)]+\frac{1}{46080n^4}[105k^{12}+k^{10}(315a^2-1935)$$
$$+k^8(525a^4-4575a^2+7075)+k^6(525a^6-5925a^4+11975a^2-939)$$
$$+k^4(315a^8-4365a^6+11045a^4-71a^2-213)$$
$$+k^2(105a^{10}-1725a^8+5275a^6+61a^4+305a^2+915)$$
$$\left.+(15a^{12}-285a^{10}+1025a^8+9a^6+63a^4+315a^2+945)]+\ldots\right\}. \tag{24}$$

In this expression we have written

$$H(x)=\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2},$$
$$G(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}u^2}\,du,$$
$$a=\frac{h-\rho k}{\sqrt{(1-\rho^2)}}\quad\text{and}\quad b=\frac{k-\rho h}{\sqrt{(1-\rho^2)}}.$$

The numerical values of H and G are conveniently obtained from the tables of the probability function given by the National Bureau of Standards (1953).

4. ASYMPTOTIC EXPRESSION FOR THE EQUI-COORDINATE PERCENTAGE POINTS

For any P $(0<P<1)$, we define a point (h,k) to be a $100P$-percentage point if it satisfies the relationship

$$P=\int_{-\infty}^{k}\int_{-\infty}^{h} g_n(u,v;\rho)\,du\,dv. \tag{25}$$

For fixed P, this defines a curve in the (h,k) plane. Of particular interest in the applications is the unique point on this curve where $h=k$; this may be called the equi-coordinate $100P$-percentage point and the common value of the co-ordinates denoted by $t=t_n(P;\rho)$.

For fixed P, the value of t can be determined for any n by trial and error using formula (10) or (11). However, this procedure becomes laborious with increasing n. In this section we shall derive an asymptotic expression in powers of n^{-1} which expresses t in terms of the corresponding quantity x for the bivariate normal distribution; the latter can be obtained by interpolation in the tables of K. Pearson (1931). This expression yields a good approximation to t even for the moderately small values of n. The method of derivation is essentially the same as that used by Fisher (1941) in deriving an asymptotic expression for the percentage points of the univariate Student t-distribution.

By definition of t and x, we have

$$P=\int_{-\infty}^{t}\int_{-\infty}^{t} g_n(u,v;\rho)\,du\,dv=\int_{-\infty}^{x}\int_{-\infty}^{x} g(u,v;\rho)\,du\,dv. \tag{26}$$

Expanding the right-hand member of (26) in a Taylor series as a function of x around $x=t$, we obtain

$$\begin{aligned}P=&\int_{-\infty}^{t}\int_{-\infty}^{t} g(u,v;\rho)\,du\,dv\\&+2(x-t)\,HG+(x-t)^2\,H\{-tG+G'\}+\tfrac{1}{3}(x-t)^3\,H\{(t^3-1)\,G-t(c^2+2)\,G'\}\\&+\tfrac{1}{12}(x-t)^4\,H\{-(t^3-3t)\,G+[t^2(c^4+3c^2+3)-(c^2+3)]\,G'\}+\ldots,\end{aligned} \tag{27}$$

where $c=\sqrt{\dfrac{1-\rho}{1+\rho}}$, $H=H(t)$, $G=G(ct)$ and $G'=\dfrac{dG(ct)}{dt}=cH(ct)$.

On the other hand, using (23) with $h=k=t$, the middle member of (26) gives the following alternative expression for P:

$$\begin{aligned}P=&\int_{-\infty}^{t}\int_{-\infty}^{t} g(u,v;\rho)\,du\,dv\\&-\frac{tH}{2n}\{(t^2+1)\,G-tG'\}-\frac{tH}{48n^2}\{(3t^6-7t^4-5t^2-3)\,G-tG'[3t^4(c^4+3c^2+3)-t^2(c^2+5)-3]\}\\&-\frac{tH}{192n^3}\{(t^{10}-11t^8+14t^6+6t^4-3t^2-15)\,G-tG'[t^8(c^8+5c^6+10c^4+10c^2+5)\\&-t^6(7c^6+29c^4+46c^2+34)-t^4(c^4+c^2-6)-t^2(5c^2+3)-15]\}\\&-\frac{tH}{46080n^4}\{(15t^{14}-375t^{12}+2225t^{10}-2141t^8-939t^6-213t^4+915t^2+945)\,G\\&-tG'[15t^{12}(c^{12}+7c^{10}+21c^8+35c^6+35c^4+21c^2+7)\\&-15t^{10}(19c^{10}+115c^8+291c^6+395c^4+305c^2+129)\\&+5t^8(205c^8+1055c^6+2209c^4+2395c^2+1415)+t^6(9c^6+61c^4-71c^2-939)\\&+t^4(63c^4+305c^2-213)+15t^2(21c^2+61)+945]\}-\ldots.\end{aligned} \tag{28}$$

Equating (27) and (28), we note that the first terms of each are identical and can be cancelled. If we equate the second terms, we can obtain a solution for $x-t$ as a function of t which is correct to terms of $O(n^{-1})$. Using the result and equating the second and third terms, we can obtain a solution correct to terms of $O(n^{-2})$. Continuing in this way, we obtain the following:

$$
\begin{aligned}
x = t &- \frac{t}{4n}\left\{t^2+1-\frac{tG'}{G}\right\} - \frac{t}{96n^2}\left\{-13t^4-8t^2-3-\frac{tG'}{G}[t^4(3c^4+9c^2)\right.\\
&+t^2(-c^2-17)-6]+\left(\frac{tG'}{G}\right)^2(-9t^2-6)+3\left.\left(\frac{tG'}{G}\right)^3\right\}\\
&-\frac{t}{384n^3}\left\{35t^6+19t^4+t^2-15-\frac{tG'}{G}[t^8(c^8+5c^6+7c^4)+t^6(-7c^6-32c^4-57c^2)\right.\\
&+t^4(-c^4-3c^2+55)+t^2(-6c^2+18)-12]+\left(\frac{tG'}{G}\right)^2[t^6(-6c^4-21c^2)\\
&+t^4(-3c^4-13c^2+50)+t^2(-2c^2+40)+9]\\
&-\left(\frac{tG'}{G}\right)^3[t^4(-3c^4-12c^2+14)+t^2(-2c^2+36)+12]\\
&+\left(\frac{tG'}{G}\right)^4[t^2(-c^2+13)+9]-3\left.\left(\frac{tG'}{G}\right)^5\right\}\\
&-\frac{t}{92160n^4}\left\{-6271t^8-3224t^6+102t^4+1680t^2+945\right.\\
&-\frac{tG'}{G}[t^{12}(15c^{12}+105c^{10}+255c^8+225c^6)\\
&+t^{10}(-285c^{10}-1785c^8-4245c^6-4185c^4)+t^8(1025c^8+5695c^6+13265c^4+17275c^2)\\
&+t^6(9c^6+241c^4+1769c^2-11319)+t^4(48c^4+1040c^2-3548)\\
&+t^2(420c^2+2820)+1800]+\left(\frac{tG'}{G}\right)^2[t^{10}(-165c^8-960c^6-1575c^4)\\
&+t^8(-60c^8+390c^6+3900c^4+11430c^2)+t^6(330c^6+2245c^4+7860c^2-13765)\\
&+t^4(120c^4+2680c^2-11380)+t^2(750c^2-1320)+1440]\\
&-\left(\frac{tG'}{G}\right)^3[t^8(-105c^8-750c^6-735c^4+2700c^2)\\
&+t^6(270c^6+2670c^4+10080c^2-7020)+t^4(385c^4+3880c^2-14735)\\
&+t^2(1140c^2-6600)-180]+\left(\frac{tG'}{G}\right)^4[t^6(-90c^6+570c^4+3750c^2-1350)\\
&+t^4(510c^4+3750c^2-9315)+t^2(870c^2-7620)-1710]\\
&-\left(\frac{tG'}{G}\right)^5[t^4(255c^4+1515c^2-2310)+t^2(525c^2-4755)-1620]\\
&\left.+\left(\frac{tG'}{G}\right)^6[t^2(150c^2-1275)-900]-\left(\frac{tG'}{G}\right)^7[-255]\right\}-\ldots. \qquad (29)
\end{aligned}
$$

Using a cumulative term-by-term procedure similar to that described above, the series (29) can be inverted to give t as a function of x. The final result may be written as follows, where G and G' now represent the same functions as in (27) with the argument changed from t to x:

$$
\begin{aligned}
t = x &+ \frac{x}{4n}\left\{x^2+1-\frac{xG'}{G}\right\} + \frac{x}{96n^2}\left\{5x^4+16x^2+3\right. \\
&-\frac{xG'}{G}[x^4(3c^4+3c^2)+x^2(-7c^2+13)+12]+\left(\frac{xG'}{G}\right)^2[x^2(-6c^2-3)+12] \\
&\left.-\left(\frac{xG'}{G}\right)^3[3]\right\}+\frac{x}{384n^3}\left\{3x^6+19x^4+17x^2-15-\frac{xG'}{G}[x^8(c^8+2c^6+c^4)\right. \\
&+x^6(-10c^6-7c^4+3c^2)+x^4(24c^4+6c^2+13)+x^2(-29c^2+43)-3] \\
&+\left(\frac{xG'}{G}\right)^2[x^6(-6c^6-9c^4-3c^2)+x^4(38c^4+23c^2-4)+x^2(-60c^2+15)+36] \\
&-\left(\frac{xG'}{G}\right)^3[x^4(12c^4+12c^2+2)+x^2(-46c^2-16)+39] \\
&\left.+\left(\frac{xG'}{G}\right)^4[x^2(-10c^2-5)+18]-\left(\frac{xG'}{G}\right)^5[3]\right\} \\
&+\frac{x}{92160n^4}\left\{79x^8+776x^6+1482x^4-1920x^2-945\right. \\
&-\frac{xG'}{G}[x^{12}(15c^{12}+45c^{10}+45c^8+15c^6)+x^{10}(-345c^{10}-705c^8-375c^6-15c^4) \\
&+x^8(2195c^8+2845c^6+775c^4+125c^2) \\
&+x^6(-5241c^6-3589c^4+379c^2+471)+x^4(5478c^4-290c^2+3032) \\
&+x^2(-1950c^2+1770)-3240]+\left(\frac{xG'}{G}\right)^2[x^{10}(-210c^{10}-525c^8-420c^6-105c^4) \\
&+x^8(2880c^8+4890c^6+1980c^4-30c^2) \\
&+x^6(-11830c^6-11445c^4-900c^2-155)+x^4(19240c^4+7600c^2+400) \\
&+x^2(-12780c^2+5130)-360]-\left(\frac{xG'}{G}\right)^3[x^8(915c^8+1830c^6+1095c^4+180c^2) \\
&+x^6(-8370c^6-10950c^4-3000c^2+60) \\
&+x^4(22835c^4+15080c^2-145)+x^2(-23280c^2-2280)+7110] \\
&+\left(\frac{xG'}{G}\right)^4[x^6(-1860c^6-2790c^4-1110c^2-90) \\
&+x^4(11280c^4+10050c^2+1395)+x^2(-18960c^2-6480)+9360] \\
&-\left(\frac{xG'}{G}\right)^5[x^4(1965c^4+1965c^2+390)+x^2(-7245c^2-3285)+5760] \\
&\left.+\left(\frac{xG'}{G}\right)^6[x^2(-1050c^2-525)+1800]-\left(\frac{xG'}{G}\right)^7[225]\right\}+\ldots. \qquad (30)
\end{aligned}
$$

5. TABLES AND NUMERICAL RESULTS

Consider the expression $P_n(h, k; \rho)$ with $h = k = t$, say, and denote the resulting expression by $P_n(t; \rho)$. This expression is tabulated for selected value of t and $n = 1\,(1)\,30\,(3)\,45\,(15)\,120$, 150, 300, 600, ∞ for $\rho = +0{\cdot}5$ in Table 1 and for $\rho = -0{\cdot}5$ in Table 2. The computation was carried out in the following manner. For each t, the exact formulae, (10) and (11), were used for $n = 1, 2$, etc., until agreement was reached in the fifth decimal place with the asymptotic formula (23). For all subsequent values of n, the asymptotic formula (23) was used and checked in specific instances by formulae (10) and (11). In all cases, the result was rounded in the usual way to the nearest digit in the fifth decimal place. By this procedure, the entries in Tables 1 and 2 should be correct to the number of figures given except in a few cases where the error may be as much as 0·0000055.

It should be noted that if $P_n(t; \rho)$ is required for t negative, it can be obtained from the corresponding expression for t positive using the relation

$$P_n(-t; \rho) = 1 + P_n(t; \rho) - 2F_n(t), \tag{31}$$

where $F_n(t)$ represents the c.d.f. of the univariate Student t-distribution with n degrees of freedom, which is tabulated, for example, by Hartley & Pearson (1950).

The value of $t_n(P; \rho)$ is tabulated for selected values of P and $n = 1\,(1)\,30\,(3)\,45\,(15)\,120$, 150, 300, 600, ∞ for $\rho = +0{\cdot}5$ in Table 3 and for $\rho = -0{\cdot}5$ in Table 4. For each P, the exact formulae (10) and (11) were used in a trial and error manner to determine the smallest value of t to three places of decimals which achieves the desired value of P, for $n = 1, 2$, etc., until agreement was reached in the third decimal place with the asymptotic formula (30). For all subsequent values of n, the asymptotic formula (30) was used and checked in specific instances by formulae (10) and (11). By this procedure, the entries in the table were always rounded to the next higher value and should be in error by at most one unit in the last figure given.

In Tables 5 and 6 we give the numerical values of the coefficients in the asymptotic expressions used in the construction of Tables 1, 2, 3 and 4. It follows from (23) and (30) that, for $h = k = t$, we can write the first five terms of the asymptotic expressions for $P_n(t; \rho)$ and $t_n(P; \rho)$ in the form

$$P_n(t; \rho) = A_0 + \frac{A_1}{n} + \frac{A_2}{n^2} + \frac{A_3}{n^3} + \frac{A_4}{n^4} \tag{32}$$

and

$$t_n(P; \rho) = B_0 + \frac{B_1}{n} + \frac{B_2}{n^2} + \frac{B_3}{n^3} + \frac{B_4}{n^4}. \tag{33}$$

For the special cases $\rho = \pm \frac{1}{2}$, Table 5 gives A_0, A_1, A_2, A_3 and A_4 for selected values of t and Table 6 gives B_0, B_1, B_2, B_3 and B_4 for selected values of P. A_0 and B_0 were obtained by interpolation in Pearson's tables (1931). The remaining coefficients were calculated using (23) and (30). For n large, it is clear that the only limit on the accuracy of the asymptotic formulae is the accuracy to which A_0 and B_0 can be determined. In the last column of these tables are given the smallest values of n for which (32) and (33) agree with the exact formulae (using the same rounding procedure in the asymptotic case as in the corresponding exact case) to the number of places given in Tables 1, 2, 3 and 4.

The authors wish to express their indebtedness to Prof. R. E. Bechhofer, under whose guidance this work was done, and also to Mrs Shirley Hockett who computed the major part of the tables.

Table 1. *Probability integral of a bivariate t-distribution with* $\rho = +0{\cdot}5$

Degrees of freedom (n)	Co-ordinate (h)												
	0·00	0·25	0·50	0·75	1·00	1·25	1·50	1·75	2·00	2·25	2·50	3·00	3·50
1	0·33333	0·41677	0·50000	0·57386	0·63497	0·68411	0·72338	0·75495	0·78063	0·80178	0·81943	0·84704	0·86754
2	·33333	·42592	·52017	·60697	·68114	·74145	·78925	·82671	·85607	·87919	·89754	·92417	·94197
3	·33333	·42955	·52818	·62036	·70018	·76550	·81702	·85680	·88720	·91038	·92812	·95238	·96723
4	·33333	·43148	·53245	·62755	·71052	·77869	·83231	·87326	·90403	·92692	·94394	·96612	·97874
5	·33333	·43267	·53510	·63202	·71701	·78702	·84196	·88362	·91449	·93707	·95344	·97397	·98493
6	0·33333	0·43348	0·53689	0·63507	0·72146	0·79275	0·84862	0·89074	0·92163	0·94387	0·95972	0·97894	0·98864
7	·33333	·43407	·53819	·63728	·72470	·79693	·95348	·89592	·92679	·94875	·96413	·98230	·99104
8	·33333	·43451	·53917	·63895	·72716	·80012	·85719	·89987	·93070	·95241	·96743	·98470	·99268
9	·33333	·43486	·53994	·64027	·72909	·80263	·86011	·90297	·93375	·95525	·96995	·98650	·99386
10	·33333	·43513	·54056	·64133	·73065	·80466	·86247	·90547	·93621	·95751	·97194	·98788	·99473
11	0·33333	0·43536	0·54107	0·64220	0·73194	0·80633	0·86442	0·90753	0·93823	0·95936	0·97355	0·98897	0·99540
12	·33333	·43555	·54150	·64292	·73301	·80773	·86605	·90926	·93991	·96089	·97487	·98985	·99593
13	·33333	·43572	·54186	·64354	·73393	·80893	·86744	·91073	·94134	·96219	·97598	·99057	
14	·33333	·43586	·54217	·64407	·73472	·80995	·86864	·91199	·94256	·96330	·97693	·99117	
15	·33333	·43598	·54243	·64453	·73540	·81085	·86968	·91309	·94362	·96425	·97774	·99168	
16	0·33333	0·43608	0·54267	0·64494	0·73600	0·81163	0·87060	0·91406	0·94455	0·96509	0·97845	0·99212	
17	·33333	·43618	·54288	·64530	·73653	·81233	·87141	·91491	·94538	·96582	·97906	·99250	
18	·33333	·43626	·54306	·64561	·73701	·81295	·87213	·91567	·94611	·96648	·97961	·99282	
19	·33333	·43633	·54323	·64590	·73743	·81350	·87277	·91635	·94676	·96706	·98009	·99311	
20	·33333	·43640	·54338	·64616	·73781	·81401	·87336	·91696	·94735	·96758	·98053	·99337	
21	0·33333	0·43646	0·54351	0·64639	0·73816	0·81446	0·87389	0·91752	0·94788	0·96805	0·98092	0·99360	
22	·33333	·43652	·54364	·64660	·73847	·81487	·87438	·91803	·94837	·96848	·98127	·99380	
23	·33333	·43657	·54375	·64680	·73876	·81525	·87482	·91849	·94881	·96887	·98159	·99399	
24	·33333	·43661	·54385	·64697	·73903	·81560	·87522	·91891	·94922	·96923	·98189	·99415	
25	·33333	·43666	·54395	·64714	·73927	·81592	·87560	·91930	·94959	·96956	·98216	·99430	
26	0·33333	0·43669	0·54404	0·64729	0·73950	0·81621	0·87594	0·91967	0·94993	0·96986	0·98240	0·99444	
27	·33333	·43673	·54412	·64743	·73971	·81649	·87626	·92000	·95025	·97014	·98263	·99457	
28	·33333	·43676	·54419	·64756	·73990	·81674	·87656	·92031	·95055	·97040	·98284	·99469	
29	·33333	·43680	·54426	·64768	·74008	·81698	·87683	·92060	·95083	·97064	·98304	·99479	
30	·33333	·43683	·54433	·64779	·74025	·81720	·87709	·92087	·95108	·97086	·98322	·99489	

30	·33333	·43683	·54433	·64779	·74025	·81720	·87709	·92087	·95108	·97086	·98322	·99489
33	0·33333	0·43690	0·54450	0·64809	0·74070	0·81779	0·87778	0·92159	0·95176	0·97146	0·98370	0·99515
36	·33333	·43697	·54465	·64834	·74107	·81828	·87835	·92219	·95233	·97195	·98409	·99536
39	·33333	·43702	·54477	·64855	·74138	·81869	·87883	·92269	·95281	·97236	·98443	·99554
42	·33333	·43707	·54487	·64873	·74165	·81905	·87925	·92313	·95322	·97272	·98471	·99568
45	·33333	·43711	·54497	·64889	·74189	·81936	·87961	·92350	·95358	·97303	·98495	·99581
48	0·33333	0·43715	0·54505	0·64903	0·74210	0·81963	0·87992	0·92383	0·95389	0·97329	0·98517	0·99592
51	·33333	·43718	·54512	·64915	·74228	·81987	·88020	·92412	·95416	·97353	·98535	·99601
54	·33333	·43721	·54518	·64926	·74244	·82008	·88045	·92438	·95441	·97374	·98552	·99609
57	·33333	·43723	·54523	·64935	·74258	·82027	·88067	·92462	·95462	·97393	·98567	·99617
60	·33333	·43725	·54529	·64944	·74271	·82044	·88087	·92482	·95482	·97410	·98580	·99623
75	0·33333	0·43734	0·54548	0·64977	0·74321	0·82110	0·88164	0·92562	0·95557	0·97473	0·98630	0·99648
90	·33333	·43740	·54560	·64999	·74354	·82153	·88215	·92615	·95607	·97516	·98663	·99664
105	·33333	·43744	·54570	·65015	·74378	·82185	·88251	·92653	·95642	·97546	·98687	·99675
120	·33333	·43747	·54576	·65027	·74396	·82208	·88279	·92682	·95669	·97569	·98704	·99683
150	0·33333	0·43751	0·54586	0·65044	0·74420	0·82241	0·88317	0·92721	0·95706	0·97600	0·98729	0·99695
300	·33333	·43760	·54605	·65077	·74470	·82307	·88394	·92801	·95781	·97663	·98777	·99717
600	·33333	·43764	·54615	·65094	·74495	·82340	·88432	·92841	·95818	·97695	·98801	·99728
∞	·33333	·43768	·54625	·65110	·74520	·82373	·88471	·92881	·95855	·97726	·98825	·99738

Degrees of freedom (n)	Co-ordinate (h)												
	4·0	4·5	5·0	5·5	6·0	6·5	7·0	7·5	8·0	8·5	9·0	9·5	10·0
1	0·88333	0·89580	0·90590	0·91428	0·92124	0·92719	0·93231	0·93676	0·94066	0·94412	0·94719	0·94995	0·95243
2	·95433	·96321	·96978	·97476	·97862	·98166	·98411	·98610	·98774	·98911	·99026		
3	·97667	·98289	·98712	·99009									
4	·98616	·99068											
5	·99095												
6	0·99365												

Note. The relationship between h, P and n, where P is the quantity given in the body of the table, is given by

$$P = \int_{-\infty}^{h}\int_{-\infty}^{h} \frac{1}{\pi\sqrt{3}} \left\{1 + \frac{4(x^2 - xy + y^2)}{3n}\right\}^{-\frac{1}{2}(n+2)} dx\,dy.$$

Table 2. *Probability integral of a bivariate t-distribution with* $\rho = -0{\cdot}5$

Degrees of freedom (n)	Co-ordinate (h)												
	0·00	0·25	0·50	0·75	1·00	1·25	1·50	1·75	2·00	2·25	2·50	3·00	3·50
1	0·16667	0·26044	0·36554	0·46120	0·54022	0·60327	0·65333	0·69336	0·71332	0·75247	0·77465	0·80928	0·83495
2	·16667	·26993	·38889	·50259	·59968	·67778	·73892	·78635	·82320	·85262	·87479	·90763	·92947
3	·16667	·27368	·39817	·51957	·62479	·70974	·77566	·82580	·86365	·89222	·91391	·94330	·96114
4	·16667	·27568	·40313	·52878	·63863	·72750	·79605	·84747	·88549	·91356	·93396	·96043	·97531
5	·16667	·27691	·40620	·53455	·64740	·73881	·80900	·86111	·89902	·92631	·94588	·97007	·98280
6	0·16667	0·27774	0·40829	0·53851	0·65345	0·74663	0·81799	0·87046	0·90820	0·93492	0·95368	0·97609	0·98721
7	·16667	·27835	·40980	·54139	·65788	·75237	·82449	·87727	·91482	·94103	·95914	·98010	·99002
8	·16667	·27880	·41095	·54359	·66126	·75675	·82949	·88245	·91981	·94559	·96314	·98295	·99193
9	·16667	·27916	·41185	·54531	·66393	·76021	·83342	·88651	·92369	·94911	·96619	·98506	·99328
10	·16667	·27945	·41257	·54670	·66608	·76302	·83661	·88978	·92681	·95191	·96859	·98667	·99428
11	0·16667	0·27968	·041316	0·54784	0·66786	0·76533	0·83923	0·89247	0·92935	0·95418	0·97052	0·98793	0·99503
12	·16667	·27988	·41366	·54880	·66936	·76727	·84144	·89472	·93148	·95606	·97211	·98894	·99562
13	·16667	·28005	·41408	·54962	·67063	·76893	·84331	·89663	·93327	·95763	·97343	·98976	·99609
14	·16667	·28019	·41444	·55033	·67172	·77036	·84493	·89827	·93481	·95898	·97455	·99045	
15	·16667	·28031	·41476	·55093	·67268	·77160	·84633	·89970	·93614	·96014	·97550	·99103	
16	0·16667	0·28042	0·41503	0·55147	0·67352	0·77269	0·84757	0·90095	0·93730	0·96115	0·97633	0·99152	
17	·16667	·28052	·41528	·55194	·67426	·77366	·84866	·90205	·93832	·96204	·97706	·99195	
18	·16667	·28061	·41549	·55236	·67492	·77452	·84963	·90304	·93924	·96282	·97770	·99232	
19	·16667	·28068	·41569	·55274	·67551	·77529	·85051	·90392	·94005	·96352	·97826	·99264	
20	·16667	·28075	·41586	·55308	·67605	·77599	·85129	·90471	·94078	·96415	·97877	·99292	
21	0·16667	0·28081	0·41602	0·55339	0·67654	0·77663	0·85201	0·90543	0·94144	0·96472	0·97923	0·99318	
22	·16667	·28087	·41617	·55367	·67698	·77720	·85266	·90608	·94204	·96523	·97964	·99341	
23	·16667	·28092	·41630	·55393	·67738	·77773	·85325	·90668	·94259	·96569	·98001	·99361	
24	·16667	·28097	·41642	·55417	·67775	·77822	·85380	·90723	·94309	·96612	·98035	·99379	
25	·16667	·28102	·41653	·55438	·67810	·77866	·85430	·90774	·94356	·96651	·98066	·99396	
26	0·16667	0·28106	0·41663	0·55458	0·67841	0·77908	0·85477	0·90820	0·94398	0·96687	0·98094	0·99411	
27	·16667	·28109	·41673	·55477	·67871	·77946	·85520	·90863	·94438	·96721	·98121	·99425	
28	·16667	·28113	·41682	·55494	·67898	·77982	·85560	·90903	·94474	·96751	·98145	·99438	
29	·16667	·28116	·41690	·55511	·67923	·78015	·85597	·90941	·94508	·96780	·98168	·99450	
30	·16667	·28119	·41698	·55526	·67947	·78046	·85632	·90976	·94540	·96807	·98189	·99461	
33	0·16667	0·28127	0·41718	0·55565	0·68010								

29	·16667	·28116	·41690	·55511	·67923	·78015	·85857	·90976	·94540	·96807	·98189	·99461
30	·16667	·28119	·41698	·55526	·67947	·78046	·85632					
33	0·16667	0·28127	0·41718	0·55565	0·68010	0·78128	0·85724	0·91068	0·94623	0·96877	0·98243	0·99489
36	·16667	·28134	·41735	·55599	·68062	·78196	·85801	·91144	·94693	·96935	·98289	·99512
39	·16667	·28140	·41749	·55627	·68107	·78254	·85866	·91209	·94752	·96984	·98327	·99531
42	·16667	·28144	·41761	·55651	·68145	·78304	·85922	·91265	·94802	·97026	·98359	·99547
45	·16667	·28149	·41772	·55672	·68178	·78347	·85970	·91313	·94846	·97062	·98387	·99561
48	0·16667	0·28152	0·41781	0·55690	0·68207	0·78385	0·86013	0·91356	0·94884	0·97093	0·98411	0·99573
51	·16667	·28156	·41790	·55707	·68233	·78419	·86051	·91393	·94917	·97121	·98432	·99583
54	·16667	·28158	·41797	·55721	·68256	·78448	·86084	·91426	·94947	·97146	·98451	·99592
57	·16667	·28161	·41804	·55734	·68276	·78475	·86114	·91456	·94974	·97168	·98468	·99600
60	·16667	·28163	·41810	·55745	·68294	·78499	·86141	·91482	·94998	·97187	·98483	·99607
75	0·16667	0·28172	0·41832	0·55790	0·68365	0·78591	0·86243	0·91584	0·95089	0·97262	0·98540	0·99634
90	·16667	·28178	·41847	·55819	·68411	·78652	·86312	·91652	·95149	·97311	·98577	·99651
105	·16667	·28182	·41858	·55840	·68445	·78696	·86361	·91700	·95192	·97347	·98603	·99663
120	·16667	·28185	·41866	·55856	·68470	·78729	·86398	·91736	·95225	·97373	·98623	·99672
150	0·16667	0·28190	0·41877	0·55879	0·68505	0·78775	0·86449	0·91787	0·95270	0·97410	0·98650	0·99684
300	·16667	·28199	·41900	·55923	·68576	·78867	·86552	·91889	·95360	·97483	·98705	·99707
600	·16667	·28203	·41911	·55946	·68612	·78914	·86604	·91940	·95405	·97519	·98731	·99719
∞	·16667	·28208	·41922	·55968	·68647	·78960	·86656	·91991	·95450	·97555	·98758	·99730

Degrees of freedom (n)	Co-ordinate (h)											
	4·0	4·5	5·0	5·5	6·0	6·5	7·0	7·5	8·0	8·5	9·0	9·5	10·0
1	0·85466	0·87024	0·88284	0·89324	0·90196	0·90937	0·91575	0·92129	0·92616	0·93046	0·93429	0·93772	0·94081
2	·94458	·95540	·96339	·96944	·97412	·97782	·98078	·98319	·98518	·98684	0·98823	·98942	·99044
3	·97241	·97981	·98483	·98834	·99086								
4	·98400	·98926	·99256										
5	·98971	·99363											
6	0·99290												

Note. The relationship between h, P and n, where P is the quantity given in the body of the table, is given by

$$P = \int_{-\infty}^{h}\int_{-\infty}^{h} \frac{1}{\pi\sqrt{3}} \left\{1 + \frac{4(x^2+xy+y^2)}{3n}\right\}^{-\frac{1}{2}(n+2)} dx\,dy.$$

Table 3. *Percentage points of a bivariate t-distribution with $\rho = +0{\cdot}5$*

Degrees of freedom (n)	Probability (P)				
	0·50	0·75	0·90	0·95	0·99
1	0·500	1·708	4·696	9·511	47·733
2	·446	1·291	2·539	3·805	8·879
3	·428	1·186	2·130	2·939	5·483
4	·420	1·138	1·963	2·611	4·408
5	·414	1·111	1·874	2·441	3·900
6	0·411	1·094	1·817	2·337	3·608
7	·409	1·082	1·779	2·268	3·419
8	·407	1·073	1·751	2·218	3·287
9	·405	1·066	1·730	2·180	3·190
10	·404	1·061	1·714	2·151	3·115
11	0·403	1·056	1·701	2·128	3·056
12	·402	1·053	1·690	2·109	3·009
13	·402	1·050	1·680	2·093	2·969
14	·401	1·047	1·673	2·079	2·936
15	·401	1·045	1·666	2·067	2·908
16	0·400	1·043	1·660	2·057	2·884
17	·400	1·041	1·655	2·049	2·863
18	·400	1·040	1·651	2·041	2·844
19	·399	1·038	1·647	2·034	2·828
20	·399	1·037	1·643	2·028	2·813
21	0·399	1·036	1·640	2·022	2·800
22	·399	1·035	1·637	2·017	2·788
23	·398	1·034	1·634	2·013	2·778
24	·398	1·033	1·632	2·009	2·768
25	·398	1·033	1·629	2·005	2·759
26	0·398	1·032	1·627	2·001	2·751
27	·398	1·031	1·626	1·998	2·743
28	·398	1·031	1·624	1·995	2·736
29	·397	1·030	1·622	1·992	2·730
30	·397	1·029	1·621	1·990	2·724
33	0·397	1·028	1·617	1·983	2·708
36	·397	1·027	1·613	1·977	2·695
39	·397	1·026	1·610	1·972	2·684
42	·396	1·025	1·608	1·968	2·674
45	·396	1·024	1·606	1·965	2·666
48	0·396	1·024	1·604	1·962	2·659
51	·396	1·023	1·603	1·959	2·653
54	·396	1·023	1·601	1·957	2·648
57	·396	1·022	1·600	1·954	2·643
60	·396	1·022	1·599	1·953	2·639
75	0·395	1·020	1·594	1·945	2·622
90	·395	1·019	1·592	1·941	2·611
105	·395	1·019	1·590	1·937	2·604
120	·395	1·018	1·588	1·935	2·598
150	0·395	1·017	1·586	1·931	2·590
300	·394	1·016	1·582	1·924	2·574
600	·394	1·015	1·580	1·920	2·566
∞	0·39351	1·01391	1·57700	1·91634	2·55781

Note. The relationship between h, P and n, where h is the quantity given in the body of the table and represents the common value of the co-ordinates of the equi-coordinate percentage point (h, h), is given by

$$P = \int_{-\infty}^{h}\int_{-\infty}^{h} \frac{1}{\pi\sqrt{3}} \left\{1 + \frac{4(x^2 - xy + y^2)}{3n}\right\}^{-\frac{1}{2}(n+2)} dx\,dy.$$

Table 4. *Percentage points of a bivariate t-distribution with* $\rho = -0{\cdot}5$

Degrees of freedom (n)	Probability (P)				
	0·50	0·75	0·90	0·95	0·99
1	0·867	2·225	5·881	11·850	59·385
2	·744	1·553	2·864	4·231	9·777
3	·708	1·395	2·333	3·161	5·812
4	·691	1·325	2·121	2·767	4·595
5	·681	1·286	2·008	2·566	4·028
6	0·674	1·261	1·938	2·444	3·706
7	·670	1·243	1·891	2·363	3·499
8	·666	1·231	1·857	2·305	3·355
9	·664	1·221	1·831	2·261	3·250
10	·662	1·213	1·811	2·228	3·169
11	0·660	1·207	1·794	2·201	3·106
12	·658	1·202	1·781	2·179	3·055
13	·657	1·197	1·770	2·160	3·012
14	·656	1·193	1·760	2·145	2·977
15	·655	1·190	1·752	2·131	2·947
16	0·654	1·187	1·745	2·120	2·921
17	·654	1·185	1·739	2·110	2·898
18	·653	1·183	1·733	2·101	2·879
19	·653	1·181	1·729	2·093	2·861
20	·652	1·179	1·724	2·086	2·846
21	0·652	1·178	1·720	2·080	2·832
22	·651	1·176	1·717	2·074	2·819
23	·651	1·175	1·714	2·069	2·808
24	·651	1·174	1·711	2·064	2·797
25	·650	1·173	1·708	2·060	2·788
26	0·650	1·172	1·705	2·056	2·779
27	·650	1·171	1·703	2·052	2·771
28	·649	1·170	1·701	2·049	2·764
29	·649	1·169	1·699	2·046	2·757
30	·649	1·168	1·697	2·043	2·750
33	0·648	1·166	1·692	2·035	2·734
36	·648	1·165	1·688	2·028	2·720
39	·648	1·163	1·685	2·023	2·708
42	·647	1·162	1·682	2·018	2·698
45	·647	1·161	1·679	2·014	2·690
48	0·647	1·160	1·677	2·011	2·683
51	·646	1·159	1·675	2·008	2·676
54	·646	1·159	1·674	2·005	2·670
57	·646	1·158	1·672	2·003	2·665
60	·646	1·157	1·671	2·001	2·661
75	0·645	1·155	1·666	1·993	2·643
90	·645	1·154	1·662	1·987	2·632
105	·645	1·153	1·660	1·983	2·624
120	·644	1·152	1·658	1·980	2·618
150	0·644	1·151	1·655	1·976	2·609
300	·643	1·149	1·650	1·968	2·593
600	·643	1·148	1·648	1·964	2·584
∞	0·64235	1·14653	1·64457	1·95993	2·57568

Note. The relationship between h, P and n, where h is the quantity given in the body of the table and represents the common value of the co-ordinates of the equi-coordinate percentage point (h, h), is given by

$$P = \int_{-\infty}^{h}\int_{-\infty}^{h} \frac{1}{\pi\sqrt{3}} \left\{1 + \frac{4(x^2+xy+y^2)}{3n}\right\}^{-\frac{1}{2}(n+2)} dx\,dy.$$

Table 5. *Coefficients in the asymptotic expansion of* $P_n(t;\rho)$ *for selected values of* t *and* ρ

ρ	t	A_0	A_1	A_2	A_3	A_4	n
	0·25	0·437684	−0·025870	0·003371	0·003816	−0·001050	4
	0·50	0·546245	−0·057784	0·008999	0·006868	−0·002155	4
	0·75	0·651102	−0·100016	0·021983	0·006891	−0·001879	6
	1·00	0·745203	−0·150182	0·047374	−0·006835	0·007991	5
	1·25	0·823729	−0·198378	0·079687	−0·033130	0·036817	6
	1·50	0·884709	−0·231628	0·096254	−0·038696	0·032808	6
+0·5							
	1·75	0·928814	−0·240531	0·067469	0·052274	−0·191482	9
	2·00	0·958553	−0·223682	−0·020268	0·293449	−0·819219	12
	2·25	0·977260	−0·187525	−0·149011	0·623867	−1·618705	13
	2·50	0·988250	−0·142571	−0·276255	0·858993	−1·765249	12
	3·00	0·997382	−0·062685	−0·376815	0·432592	2·236773	18
	0·25	0·282077	−0·026677	0·003458	0·003805	−0·001048	4
	0·50	0·419223	−0·067860	0·012805	0·005089	−0·001195	4
	0·75	0·559679	−0·134168	0·044880	−0·012987	0·016283	5
	1·00	0·686472	−0·213243	0·098390	−0·052516	0·044512	7
	1·25	0·789599	−0·278618	0·131625	−0·029995	−0·127090	9
	1·50	0·866558	−0·310774	0·105272	0·112851	−0·498494	11
−0·5							
	1·75	0·919908	−0·305392	0·014720	0·322909	−0·642101	8
	2·00	0·954503	−0·269699	−0·120697	0·556270	−0·697258	9
	2·25	0·975551	−0·216425	−0·268215	0·770582	−0·926531	8
	2·50	0·987581	−0·158846	−0·387429	0·840638	−0·678115	8
	3·00	0·997300	−0·066478	−0·435425	0·171531	3·163945	17

Table 6. *Coefficients in the asymptotic expansion of* $t_n(P;\rho)$ *for selected values of* P *and* ρ

ρ	P	B_0	B_1	B_2	B_3	B_4	n
	0·50	1·000000	0·251277	0·039480	−0·030211	−0·008417	2
	0·75	1·000000	0·438753	0·193001	0·038173	−0·007468	3
+0·5	0·90	1·000000	0·798448	0·623867	0·356845	0·118483	3
	0·95	1·000000	1·098969	1·140833	0·900948	0·472589	5
	0·99	1·000000	1·832368	3·032947	3·802956	3·464277	8
	0·50	1·000000	0·284231	0·068289	−0·009598	−0·009109	2
	0·75	1·000000	0·550243	0·263130	0·102643	0·005626	2
−0·5	0·90	1·000000	0·921220	0·810736	0·441786	0·418648	4
	0·95	1·000000	1·209261	1·417233	1·120851	0·582289	6
	0·99	1·000000	1·908503	3·427819	4·670259	4·009603	8

REFERENCES

BECHHOFER, R. E., DUNNETT, C. W. & SOBEL, M. (1954). A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika*, **41**, 170.

FISHER, R. A. (1925). Expansion of Student's integral in power of n^{-1}. *Metron*, **5**, 109.

FISHER, R. A. (1941). The asymptotic approach to Behren's integral with further tables for the d-test of significance. *Ann. Eugen., Lond.*, **11**, 141.

HARTLEY, H. O. & PEARSON, E. S. (1950). Tables of the probability integral of the t-distribution. *Biometrika*, **37**, 168.

NATIONAL BUREAU OF STANDARDS (1953). *Tables of Normal Probability Functions, Applied Mathematics Series*, **23**. Washington, D.C.: Government Printing Office.

PAULSON, E. (1952). On the comparison of several experimental categories with a control. *Ann. Math. Statist.* **23**, 239.

PEARSON, K. (1931). *Tables for Statisticians and Biometricians*, Part II. London: Cambridge University Press for the Biometrika Trust.

A TWO-SAMPLE MULTIPLE DECISION PROCEDURE FOR RANKING MEANS OF NORMAL POPULATIONS WITH A COMMON UNKNOWN VARIANCE†

By ROBERT E. BECHHOFER, CHARLES W. DUNNETT‡ and MILTON SOBEL

Cornell University

A multiple decision approach to the problem of ranking populations according to their population means has been formulated by Bechhofer (1954). A single-sample solution to this problem was presented by him for the case of normal populations with known variances. In the present paper the case of normal populations with unknown equal variances is considered. A two-sample procedure is proposed as a solution to the latter problem; this procedure is of the type used by Stein (1945) for obtaining a test of Student's hypothesis with power independent of the variance. Tables are available which enable the experimenter to apply this procedure with little computational effort to the ranking of two or three populations with unknown but equal variances. The same tables can be used when the variance ratios are arbitrary and known. Graphic comparisons of the expected sample size are made for the single-sample and two-sample procedures.

1. Introduction

Because of the widespread popularity of the analysis of variance as a statistical technique for analysing experimental data, there sometimes is a strong tendency to misapply it. Many problems are forced into the mould of a test of homogeneity (i.e. a test of the *equality* of population means) when actually such a test will not answer the basic problem of the experimenter. Moreover, there are many practical situations in which the experimenter actually knows *a priori* that the population means are unequal. In these situations he often desires either a complete or partial ranking of the populations according to their population means, e.g. he may wish to find the population with the largest mean. Thus from the statistical point of view, the experimenter requires a *multiple decision procedure* rather than a test of homogeneity. The procedure must provide a rule which tells him how to effect his desired ranking of the populations, the rule having the property that the probability of a correct ranking will in some sense be controlled. A single-sample procedure has been proposed by Bechhofer (1954) as a solution to this problem for the case when the population variances are *known*. This paper proposes a two-sample procedure for the case when the populations have a common *unknown* variance. Except for the assumption of known variances made in the first paper, all assumptions made in both papers are the same as those for the linear model of the analysis of variance.

2. Assumptions

Let X_{ij} be normally and independently distributed chance variables from population π_i with mean μ_i and variance $\sigma_i^2 = a_i\sigma^2$ ($i = 1, 2, \ldots, k$; $j = 1, 2, \ldots$, ad inf.). We assume that σ^2 and the μ_i are unknown, and that the a_i are known positive rational numbers which without loss of generality can be taken to be integers. (Although the basic problem of interest is the case $a_i = 1$ ($i = 1, 2, \ldots, k$), it will be shown that with little extra effort we can consider the more general case of known variance ratios described above.)

† The research was sponsored by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command.

‡ Now at Lederle Laboratories, Pearl River, New York.

We denote the ranked μ_i by $\mu_{[1]} \leqslant \mu_{[2]} \leqslant \ldots \leqslant \mu_{[k]}$, (1)

and the differences between the ranked μ_i by

$$\delta_{i,j} = \mu_{[i]} - \mu_{[j]} \quad (i,j = 1,2,\ldots,k). \tag{2}$$

It is not known which population is associated with $\mu_{[i]}$.

3. The experimenter's goal and requirements

We assume that the experimenter wishes to effect either a partial or a complete ranking of the populations according to their population means, and that he specifies his goal exactly before experimentation starts. There are many goals of practical interest, but the following two are of special importance and will be treated here in detail:

Goal I: To select the population having the largest population mean.

Goal II: To select the populations having the largest, second largest, ..., and smallest population means.

Clearly the goal of selecting the population having the smallest population mean is equivalent to Goal I. For a more general goal which includes I and II, above, the reader is referred to Bechhofer (1954).

For Goal I it is assumed that the experimenter can specify the smallest value of $\delta_{k,k-1}$, say δ^*, that he is interested in detecting. The experimenter also must specify the smallest acceptable value, say P, for the probability of achieving Goal I when $\delta_{k,k-1} \geqq \delta^*$.

For Goal II it is assumed that the experimenter can specify the smallest values of the $\delta_{i+1,i}$, say $\delta^*_{i+1,i}$ $(i = 1,2,\ldots,k-1)$, that he is interested in jointly detecting. He also must specify the smallest acceptable value, say P, for the probability of achieving Goal II when $\delta_{i+1,i} \geqq \delta^*_{i+1,i}$ $(i = 1,2,\ldots,k-1)$.

4. The procedure

It should be noted that a single-sample procedure cannot satisfy the requirements of the problem, i.e. for Goal I, for example, such a procedure cannot guarantee the probability P of a correct ranking when $\delta_{k,k-1} \geqq \delta^*$, *irrespective of the true values of the unknown variances* σ_i^2. Roughly speaking, this is so because for any reasonable procedure which calls for taking a single sample of N_i observations from π_i $(i = 1,2,\ldots,k)$, the probability of a correct ranking will depend on the σ_i^2. If the true values of the σ_i^2 are sufficiently large, then it is intuitively clear that the probability of a correct ranking will be arbitrarily close to $1/k$ for Goal I and $1/k!$ for Goal II. A rigorous treatment of this point would proceed along the lines used by Dantzig (1940).

The following two-sample procedure which is of the type used by Stein (1945) will be shown to satisfy the requirements of the experimenter:

Procedure

1. Take a first sample of $a_i N_0$ observations from π_i $(i = 1,2,\ldots,k)$. (Recall that the a_i $(i = 1,2,\ldots,k)$ are known integers. Any integer N_0, for which n (defined in (4) below) is positive, will satisfy the requirements of the problem; the question of an optimum choice of N_0 will be discussed in § 5.)

2. Calculate† $$S_0^2 = \frac{1}{n}\sum_{i=1}^{k}\frac{1}{a_i}\sum_{j=1}^{a_i N_0}\left(X_{ij} - \frac{\sum_{j=1}^{a_i N_0} X_{ij}}{a_i N_0}\right)^2, \tag{3}$$

† For simplicity of notation no attempt will be made in this paper to distinguish between chance variables and their observed values.

which is an unbiased estimate of σ^2 having

$$n = N_0 \sum_{i=1}^{k} a_i - k \tag{4}$$

degrees of freedom.

3. Take a second sample of $(N - N_0)\,a_i$ observations from π_i $(i = 1, 2, \ldots, k)$, where

$$N = \max\{N_0, [2S_0^2(h_n/\delta^*)^2]\}, \tag{5}$$

the symbol $[y]$ denoting the smallest integer equal to or greater than y; here $h_n = h$ (say) is a positive constant (depending on n, P and the goal), which will be defined below.

4. Calculate for each π_i the over-all sample mean

$$\bar{X}_i = \frac{1}{a_i N} \sum_{j=1}^{a_i N} X_{ij} \quad (i = 1, 2, \ldots, k). \tag{6}$$

Denote the ranked† values of $\bar{X}_i$ by

$$\bar{X}_{[1]} < \bar{X}_{[2]} < \ldots < \bar{X}_{[k]}. \tag{7}$$

5. Rank the π_i according to the ranking of the observed $\bar{X}_i$ i.e. for Goal I select the population which gave rise to $\bar{X}_{[k]}$ as the population having the largest population mean, while for Goal II select the populations which gave rise to $\bar{X}_{[k]}, \bar{X}_{[k-1]}, \ldots,$ and $\bar{X}_{[1]}$ as the populations having the largest, second largest, ..., and smallest population mean, respectively.

Now we shall show that this procedure, by a proper choice of h, satisfies the requirements of the problem. Since the probability of achieving the goal is a strictly increasing function of $\delta_{i+1,i}$ $(i = 1, 2, \ldots, k-1)$, the requirements of the problem will be met if we can guarantee the specificed probability P of a correct ranking for Goal I when

$$\delta_{k,k-1} = \delta^* \quad \text{and} \quad \delta_{i+1,i} = 0 \quad (i = 1, 2, \ldots, k-2) \tag{8}$$

and for Goal II when $\qquad \delta_{i+1,i} = \delta^*_{i+1,i} \quad (i = 1, 2, \ldots, k-1). \tag{9}$

First we consider Goal I. Define $\bar{X}_{(i)}$ as that one of the sample means defined in (6) which comes from the population having mean $\mu_{[i]}$, i.e. the expectation of $\bar{X}_{(i)}$ is $\mu_{[i]}$ $(i = 1, 2, \ldots, k)$.

Then we can write the probability of achieving Goal I as

$$\Pr\{\bar{X}_{(i)} < \bar{X}_{(k)} \quad (i = 1, 2, \ldots, k-1)\}$$

$$= \Pr\left\{\sqrt{\frac{N}{2}}\,\frac{(\bar{X}_{(i)} - \bar{X}_{(k)} + \delta^*)}{S_0} < \sqrt{\frac{N}{2}}\,\frac{\delta^*}{S_0} \quad (i = 1, 2, \ldots, k-1)\right\}. \tag{10}$$

Denoting $\sqrt{\frac{N}{2}}\,(\bar{X}_{(i)} - \bar{X}_{(k)} + \delta^*)$ by z_i $(i = 1, 2, \ldots, k-1)$, and noting that $\sqrt{\frac{N}{2}}\,\frac{\delta^*}{S_0} \geqq h$ as a consequence of (5), it follows that the probability (10) is greater‡ than

$$\Pr\left\{\frac{z_i}{S_0} < h \quad (i = 1, 2, \ldots, k-1)\right\}. \tag{11}$$

† If, because of the limitations of the measuring instrument, two or more $\bar{X}_i$ are equal, they should be 'ranked' by using a randomized procedure which assigns equal probability to each ordering.

‡ It is also possible, following Stein's initial procedure, to devise a procedure which will guarantee a probability of a correct ranking *exactly equal* to a specified probability.

When (8) holds, the conditional joint distribution of the z_i given S_0 is the multivariate normal distribution with means 0, variances σ^2, and correlation matrix $\{\rho_{ij}\}$, where

$$\rho_{ij} = \begin{cases} 1 & \text{if } i=j \\ \frac{1}{2} & \text{if } i \neq j. \end{cases} \tag{12}$$

If in the joint distribution of the z_i and S_0 we make the transformation $z_i = t_i S_0$ and integrate out S_0, then (11) can be written

$$\int_{-\infty}^{h} \cdots \int_{-\infty}^{h} \frac{\Gamma[\frac{1}{2}(n+k-1)]}{\sqrt{k}\,(\frac{1}{2}n\pi)^{\frac{1}{2}(k-1)}\,\Gamma(\frac{1}{2}n)} \left\{1 + \frac{1}{n}\sum_{i=1}^{k-1}\sum_{j=1}^{k-1} b_{ij} t_i t_j\right\}^{-\frac{1}{2}(n+k-1)} dt_1 dt_2 \ldots dt_{k-1}, \tag{13}$$

where for Goal I we have

$$b_{ij} = \begin{cases} 2(k-1)/k & \text{if } i=j, \\ -2/k & \text{if } i \neq j. \end{cases} \tag{14}$$

Thus in order to satisfy the requirements of the problem, it is sufficient to set (13) equal to the specified probability P and solve for h.

To solve for h, tables of (13) are required. For $k = 2$, (13) reduces to the univariate Student's t-distribution which is tabulated, for example, by Merrington (1942). For $k = 3$, (13) can be regarded as a bivariate generalization of Student's t-distribution with correlation *plus* one-half; the required table has been computed and is given as Table 3 in the paper immediately preceded by Dunnett & Sobel (1954). (This table gives h to three decimal places for $P = 0{\cdot}50$, $0{\cdot}75$, $0{\cdot}90$, $0{\cdot}95$ and $0{\cdot}99$ and $n = 1\,(1)\,30\,(3)\,60\,(15)\,120$, 150, 300, 600, ∞.) For $k \geqq 4$, (13) can be regarded as a multivariate generalization of Student's t-distribution; the required integral has not been tabulated.

Now we consider Goal II. For simplicity we shall assume (see (9)) that $\delta^*_{i+1,i} = \delta^*$ $(i = 1, 2, \ldots, k-1)$. A similar argument to that given for Goal I will show that the probability of achieving Goal II is greater than the same expression (13) where for Goal II we have

$$b_{ij} = 2\min(i,j)\left\{1 - \frac{1}{k}\max(i,j)\right\}. \tag{15}$$

Again we determine the constant h to guarantee the specified probability P by referring to appropriate tables. For $k = 2$, tables of Student's t-distribution are again applicable. For $k = 3$, (13) now can be regarded as a bivariate generalization of Student's t-distribution with correlation *minus* one-half; the required table has been computed and is given by Dunnett & Sobel (1954) as Table 4. (For this table h, P and n are given as in Table 3.) For $k \geqq 4$ the required integral has not been tabulated.

Explicit formulae for the exact (for any n) and asymptotic (for large n) evaluation of (13) for $k = 3$ and *any* correlation coefficient also are given by Dunnett & Sobel (1954).

5. EXPECTED SAMPLE SIZE

Stein (1945) has given an exact formula as well as upper and lower bounds for the expected total sample size in his two-sample procedure. His results are applicable to our procedure. The expected sample size for population π_i is $a_i E(N)$, and $E(N)$ is given (to within a quantity less than unity) by the equation

$$E(N) = N_0 \Pr\left\{\chi^2_n < \frac{nN_0}{2h^2}\left(\frac{\delta^*}{\sigma}\right)^2\right\} + 2h^2\left(\frac{\sigma}{\delta^*}\right)^2 \Pr\left\{\chi^2_{n+2} > \frac{nN_0}{2h^2}\left(\frac{\delta^*}{\sigma}\right)^2\right\} + \theta \Pr\left\{\chi^2_n > \frac{nN_0}{2h^2}\left(\frac{\delta^*}{\sigma}\right)^2\right\}, \tag{16}$$

where $0 \leqq \theta \leqq 1$ and χ^2_n is the χ^2 variable with n degrees of freedom.

A numerical investigation of the relationship between $E(N)$ and N_0 was carried out for Goal I with $k = 3$, for $P = 0{\cdot}95$ and $\delta^*/\sigma = 0{\cdot}1\,(0{\cdot}1)\,0{\cdot}5$. The curves in Fig. 1 illustrate the results. They were computed using (16) and an asymptotic formula derived by Goldberg & Levine (1946).

If the value of σ were known, it would be possible to determine an $\tilde{N}$ such that a single-sample procedure requiring a sample of size $\tilde{N}a_i$ from population π_i would satisfy the requirements of the problem. Tables have been given by Bechhofer (1954) which give $\sqrt{\tilde{N}}\,(\delta^*/\sigma)$ as a function of P. For each curve in Fig. 1 there is shown for comparison a horizontal line to represent the corresponding value of $\tilde{N}$ for the single-sample procedure. For

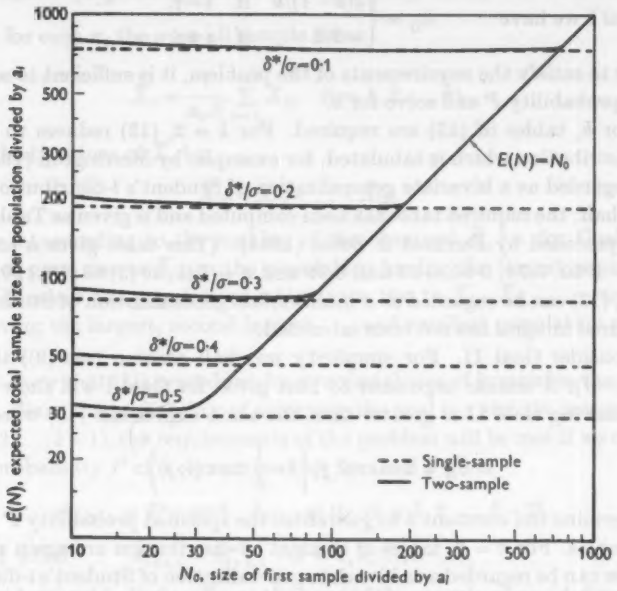


Fig. 1. Graph of expected total sample size per population for the single-sample and two-sample procedures (Goal I, $k = 3$, $P = 0{\cdot}95$).

any given P and δ^*/σ, $E(N)$ is always greater† than $\tilde{N}$, but as is shown in Fig. 1, which depicts the case $P = 0{\cdot}95$, the difference is small over a wide range of values of N_0. Each of the $E(N)$ curves in Fig. 1 appears to reach its minimum value for N_0 equal to approximately 80 % of the value of $\tilde{N}$. A feature of the distribution of N which is not evident in the diagram is the fact that its variance increases as N_0 decreases. This is an additional reason why N_0 should not be chosen much smaller than that for which the minimum value of $E(N)$ is obtained.

In practical situations where σ^2 is unknown, the experimenter sometimes can place a non-zero lower bound σ_1^2 and a finite upper bound σ_2^2 on its value. To meet the requirements of the problem, he has at least two reasonable courses of action: He can use a single-sample procedure and assume that $\sigma^2 = \sigma_2^2$. Alternatively, he can use a two-sample procedure. The curves in the diagram show that by choosing N_0 as if $\sigma^2 = \sigma_1^2$, he may be operating with $E(N)$ considerably smaller than the $\tilde{N}$ required in the single-sample procedure.

† More precisely, $E(N) > 2(h_n\sigma/\delta^*)^2 > 2(h_\infty\sigma/\delta^*)^2$ and $\tilde{N} = [2(h_\infty\sigma/\delta^*)^2]$.

6. EXAMPLES

The following examples will illustrate the procedure:

Example 1. Given three normal populations with a common *unknown* variance. Suppose that it is desired to select the population having the largest population mean and to guarantee that the probability of correctly choosing that population will be at least 0·90 when $\mu_{[3]} - \mu_{[2]} \geqq 1$. How should the experimenter proceed?

The choice of N_0, the size of the first sample from π_i $(i = 1, 2, 3)$, is optional. (Note that $a_i = 1$ $(i = 1, 2, 3)$ for this problem.) Suppose that for certain practical reasons the experimenter decides to choose $N_0 = 10$, and based on his three first samples he finds that

$$\left.\begin{aligned} &\sum_{j=1}^{10} X_{1j} = 95{\cdot}46, \qquad \sum_{j=1}^{10} X_{2j} = 114{\cdot}54, \qquad \sum_{j=1}^{10} X_{3j} = 112{\cdot}01, \\ &\sum_{j=1}^{10} X_{1j}^2 = 948{\cdot}5306, \quad \sum_{j=1}^{10} X_{2j}^2 = 1337{\cdot}1139, \quad \sum_{j=1}^{10} X_{3j}^2 = 1280{\cdot}7061. \end{aligned}\right\} \tag{17}$$

Using (3) he computes $S_0^2 = 3{\cdot}2787$, his estimate of σ^2 based on $n = 27$ degrees of freedom. Now the experimenter has specified $P = 0{\cdot}90$ and $\delta^* = 1$. Entering Table 3 in Dunnett & Sobel (1954) with $n = 27$ and $P = 0{\cdot}90$ he finds that $h = 1{\cdot}626$. Using (5) he finds that the *total* sample size per population is given by

$$N = \max\{10, [2(3{\cdot}2787)(1{\cdot}626)^2]\} = \max\{10, 18\} = 18.$$

Hence $N - N_0 = 8$ additional observations must be taken from each population. Now suppose that based on his three second samples he finds that

$$\sum_{j=11}^{18} X_{1j} = 84{\cdot}50, \quad \sum_{j=11}^{18} X_{2j} = 97{\cdot}41, \quad \sum_{j=11}^{18} X_{3j} = 95{\cdot}33. \tag{18}$$

Using the ΣX_{ij} in (17) and (18), he computes the overall means

$$\bar{X}_1 = 9{\cdot}998, \quad \bar{X}_2 = 11{\cdot}775, \quad \bar{X}_3 = 11{\cdot}519, \tag{19}$$

each of which is based on 18 observations. As his last step, he selects the second population (the population yielding the largest sample mean, $\bar{X}_2 = 11{\cdot}775$) as the population having the largest population mean.

If it had happened that $2S_0^2(1{\cdot}626)^2 \leqq 10$, no second sample would be necessary, and the experimenter would base his selection on the same means computed from (17).

The procedure for handling the problem considered in the next example is not covered explicitly in the paper, but is easily justified using the same methods.

Example 2. Consider the same problem as that described in Example 1, the only difference being that now there is available to the experimenter an earlier estimate S^2 of σ^2 based on n degrees of freedom obtained from similar experiments made on these or *other* normal populations with the same variance. How should the experimenter proceed?

There is no need now for the experimenter to take any first samples, since their major purpose was to supply the estimate of σ^2. Suppose that $S^2 = 3{\cdot}0000$ and $n = 11$. Entering Table 3 in Dunnett & Sobel (1954) with $n = 11$ and $P = 0{\cdot}90$ he finds that $h = 1{\cdot}701$. Since he has the earlier estimate he needs to take only *one* sample consisting of

$$N = \{[2S^2(h/\delta^*)^2]\} = \{[2(3{\cdot}0000)(1{\cdot}701)^2]\} = 18$$

observations from each population. He computes his sample means and selects the population yielding the largest sample mean as the population having the largest population mean.

Note 1. Even if an earlier estimate of σ^2 is available, the experimenter still may want to adopt the two-sample procedure, particularly when the number of degrees of freedom associated with his estimate is small. He will do this with the idea in mind of cutting down on the expected value and variance of N. (Recall that for Example II, N is proportional to h^2 and h increases as the degrees of freedom decrease for fixed P. For this problem with $P = 0{\cdot}90$ and $n = 1$, we have $h = 4{\cdot}696$.) If he does adopt the two-sample procedure, the estimate of σ^2 obtained from the first samples should be pooled with the earlier estimate of σ^2, and the sum of the two degrees of freedom should be used for determining h.

Note 2. The user of the tables in Dunnett & Sobel (1954) is cautioned that these tables are applicable only if the variances of the sample means, on which the final ranking is to be based, are equal.

REFERENCES

Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* (in the Press for June issue).

Dantzig, G. B. (1940). On the non-existence of tests of Student's hypothesis having power functions independent of σ. *Ann. Math. Statist.* **11**, 186.

Dunnett, C. & Sobel, M. (1954). A bivariate generalization of Student's t-distribution with tables for certain special cases. *Biometrika*, **41**, 153.

Goldberg, H. & Levine, H. (1946). Approximate formulas for the percentage points and normalization of t and χ^2. *Ann. Math. Statist.* **17**, 216.

Merrington, M. (1942). Table of percentage points of the t-distribution. *Biometrika*, **32**, 300.

Stein, C. (1945). A two sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16**, 253.

INEQUALITIES FOR THE NORMAL INTEGRAL INCLUDING A NEW CONTINUED FRACTION

BY L. R. SHENTON

College of Technology, Manchester

1. INTRODUCTION

In this note we consider the related normal integral ratios

$$R(t) = \int_t^\infty g(x)\,dx/g(t),$$

sometimes called Mills's ratio, and

$$\bar{R}(t) = \int_0^t g(x)\,dx/g(t),$$

where $g(x) = \dfrac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$, and $R(t)+\bar{R}(t) = 1/[2g(t)]$. We give a new continued fraction† for $\bar{R}(t)$ which is rapidly convergent for small values of t and which incidentally provides a new set of inequalities. The rapidity of convergence is compared with a series for $\bar{R}(t)$ and with the Laplace C.F. for $R(t)$. This assessment is similar to recent work by Teichroew (1952) on the comparative rapidity of convergence of series and C.F.'s for the elementary function e^x, $\ln(1+x)$ and $\arctan x$. Lane (1944) has also considered the same sort of thing for interpolation, comparing Newton's series and Thiele's C.F. in the case of the function 2^x.

We also prove and generalize a conjecture of Birnbaum‡ (1950) that for $t \geqslant 0$,

$$R(t) < 4/[3t+\sqrt{(t^2+8)}],$$

it being shown that there are two sequences of similar inequalities, increasing and decreasing to the limiting value $R(t)$. In the Appendix we set out a brief summary of results relating to the normal integral.

2. THE CONTINUED FRACTION FOR $R(t)$

The new C.F. is given by

$$\bar{R}(t) = \frac{t}{1} - \frac{t^2}{3} + \frac{2t^2}{5} - \frac{3t^2}{7} + \frac{4t^2}{9} - \frac{5t^2}{11} + \ldots, \tag{1}$$

and is convergent for $t \geqslant 0$. This is derived from the C.F. representation of the series

$$\bar{R}(t) = \frac{t}{1} + \frac{t^3}{1.3} + \frac{t^5}{1.3.5} + \ldots, \tag{2}$$

which may be expressed in terms of the confluent hypergeometric function

$${}_1F_1(a; b; x) = 1 + \frac{a}{b}\frac{x}{1!} + \frac{a(a+1)}{b(b+1)}\frac{x^2}{2!} + \ldots. \tag{3}$$

We then use the C.F. of Gauss (1812, pp. 134–5) in its confluent form

$${}_1F_1(1; b; x) = \frac{1}{1} - \frac{x}{b} + \frac{1x}{b+1} - \frac{bx}{b+2} + \frac{2x}{b+3} - \frac{(b+1)x}{b+4} + \ldots. \tag{4}$$

(See, for example, Perron, 1913, pp. 311–14; Wall, 1948, pp. 349–55.)

† Continued fraction will be abbreviated to C.F. throughout.

‡ Birnbaum actually conjectured $(t^2-1)R^2 - 3tR + 2 > 0$.

The expression (1) now follows since $\bar{R}(t) = t_1F_1(1; \frac{3}{2}; \frac{1}{2}t^2)$. Denoting the sth convergent of (1) by $r_s = p_s/q_s$, it follows that p_s and q_s satisfy the recurrence relations

$$\left.\begin{aligned} y_{2s+1} &= (4s+1)\,y_{2s} + 2st^2 y_{2s-1} \\ y_{2s} &= (4s-1)\,y_{2s-1} - (2s-1)\,t^2 y_{2s-2} \end{aligned}\right\} \quad (s = 1, 2, \ldots), \tag{5}$$

with $p_0 = 0$, $p_1 = t$, $q_0 = 1$, $q_1 = 1$.

It may be shown that for $t > 0$, $q_{4s} > 0$, $q_{4s+1} > 0$, so that from (5) we have

$$r_0 < r_1 < r_4 < r_5 < r_8 < r_9 < \ldots < \bar{R} \quad (t > 0). \tag{6}$$

The situation for the remaining convergents is not so simple, but it transpires that

$$r_{4s+2} > r_{4s+3} > r_{4s+6} > r_{4s+7} > \ldots > \bar{R} \quad (t > 0), \tag{7}$$

for some positive integer s. We therefore have the following simple inequalities

$$r_1 = t < r_4 = \frac{105t + 5t^3}{105 - 30t^2 + 15t^4} < r_5 = \frac{945t + 105t^3 + 8t^5}{945 - 210t^2 + 15t^4} < \bar{R} \quad (t > 0), \tag{8}$$

$$r_2 = \frac{3t}{3 - t^2} > r_3 = \frac{15t + 2t^3}{15 - 3t^2} > r_6 = \frac{10395t + 630t^3 + 63t^5}{10395 - 2835t^2 + 315t^4 - 15t^6} > \bar{R}, \tag{9}$$

the ranges of t for (9) being $0 < t < \sqrt{3}$, $0 < t < \sqrt{5}$, $0 < t < \sqrt{8{\cdot}283}$ respectively.

We now write the C.F. as

$$\bar{R}(t) = \frac{t}{1} - \frac{\overset{*}{t^2}}{3} + \frac{\overset{*}{2t^2}}{5} - \frac{3t^2}{7} + \frac{4t^2}{9} - \frac{\overset{*}{5t^2}}{11} + \frac{\overset{*}{6t^2}}{13} - \frac{7t^2}{15} + \frac{8t^2}{17} - \frac{\overset{*}{9t^2}}{19} + \frac{\overset{*}{10t^2}}{21} - \ldots, \tag{10}$$

the convergents corresponding to the asterisks forming a decreasing sequence exceeding $\bar{R}(t)$, the remaining convergents forming an increasing sequence less than $\bar{R}(t)$, this state of affairs holding sooner or later for all $t > 0$. It certainly *always* holds for $0 < t < \sqrt{3}$. Otherwise, for example, if $t = 2$, $2{\cdot}5$, $3{\cdot}0$ it holds from the 3rd, 4th and 7th convergents, respectively, onwards. Further properties of the C.F. are listed in the Appendix.

Numerical illustration

We give in Table 1 the convergents for the examples $t = 0{\cdot}1$ and $t = 3{\cdot}0$. It will be seen that for $t = 0{\cdot}1$ eight convergents give accuracy in the 22nd place of decimals, a rapid rate of convergence of about 6 decimal places for every two convergents. For $t = 3{\cdot}0$ the convergence is slow at first, being due to the negative value of some of the early denominators. However, convergence soon accelerates, there being a gain of four decimal places between the 16th and 20th convergents. We may mention that the increase in the number of digits in the convergents of the C.F. creates a computational difficulty. Thus with $t = 2{\cdot}0$ the 18th convergent involves 21 digits; with $t = 2{\cdot}5$ employing an equivalence transformation to clear decimals, the 22nd convergent involves 35 digits. This situation can be avoided if we know the number of terms required, or by curtailment at the expense of some loss of accuracy. Three methods of using C.F.'s on high-speed computing machines are discussed by Teichroew (1952).

In Table 2 we give the values of the absolute differences $|\nabla r_s| = |r_s - r_{s-1}|$, and the corresponding differences $|\nabla T_s| = |T_s - T_{s-1}|$ for the series (referred to as series C)

$$\bar{R}(t) = e^{\frac{1}{2}t^2}\left(t - \frac{t^3}{3.2.1!} + \frac{t^5}{5.2^2.2!} - \ldots\right) = \lim_{s \to \infty} T_s,$$

Table 1. *Convergents of* $\bar{R}(t)$

	s	r_s
$t=0{\cdot}1$	1	0·1
	2	0·100, 334, 4*
	3	0·100, 334, 001, 3*
	4	0·100, 334, 000, 953, 2
	5	0·100, 334, 000, 953, 439, 993
	6	0·100, 334, 000, 953, 440, 116, 2*
	7	0·100, 334, 000, 953, 440, 116, 180, 63*
	8	0·100, 334, 000, 953, 440, 116, 180, 61
$t=3{\cdot}0$	1	3
	2	−1·5
	3	−8·2
	7	164·0
	8	98·7
	15	112·515, 2*
	16	112·515, 1
	19	112·515, 152, 98*
	20	112·515, 152, 96

Note. Convergents marked with an asterisk exceed the value of $\bar{R}(t)$.

Table 2. *Differences of successive convergents of* $\bar{R}(t)$ *and of corresponding partial sums of the series C*

	s	$\lvert\nabla r_s\rvert$	$\lvert\nabla T_s\rvert$		s	$\lvert\nabla r_s\rvert$	$\lvert\nabla T_s\rvert$
$t=0{\cdot}1$	2	$\bar{3}{\cdot}33$	$\bar{3}{\cdot}17$	$t=2{\cdot}0$	6	0·28	0·36
	4	$\bar{9}{\cdot}39$	$\bar{9}{\cdot}30$		8	$\bar{2}{\cdot}52$	$\bar{1}{\cdot}25$
	6	$\overline{15}{\cdot}12$	$\overline{15}{\cdot}24$		10	$\bar{4}{\cdot}70$	$\bar{2}{\cdot}11$
	8	$\overline{22}{\cdot}18$	$\overline{21}{\cdot}10$		12	$\bar{6}{\cdot}62$	$\bar{4}{\cdot}33$
$t=0{\cdot}5$	2	$\bar{1}{\cdot}45$	$\bar{1}{\cdot}24$		14	$\bar{8}{\cdot}40$	$\bar{6}{\cdot}72$
	4	$\bar{4}{\cdot}34$	$\bar{4}{\cdot}26$		16	$\overline{10}{\cdot}19$	$\bar{7}{\cdot}12$
	6	$\bar{8}{\cdot}68$	$\bar{7}{\cdot}13$		18	$\overline{13}{\cdot}7$	$\bar{9}{\cdot}16$
	8	$\overline{12}{\cdot}64$	$\overline{11}{\cdot}36$	$t=3{\cdot}0$	14	$\bar{2}{\cdot}4$	0·5
	10	$\overline{16}{\cdot}36$	$\overline{15}{\cdot}61$		16	$\bar{4}{\cdot}9$	$\bar{1}{\cdot}4$
$t=1{\cdot}0$	2	0·50	0·27		18	$\bar{5}{\cdot}2$	$\bar{2}{\cdot}3$
	4	$\bar{2}{\cdot}63$	$\bar{2}{\cdot}49$		20	$\bar{7}{\cdot}2$	$\bar{3}{\cdot}2$
	6	$\bar{4}{\cdot}21$	$\bar{4}{\cdot}39$		22	$\bar{9}{\cdot}3$	$\bar{5}{\cdot}6$
	8	$\bar{7}{\cdot}31$	$\bar{6}{\cdot}17$		24	$\overline{11}{\cdot}3$	$\bar{6}{\cdot}2$
	10	$\overline{10}{\cdot}27$	$\bar{9}{\cdot}47$		26	$\overline{13}{\cdot}2$	$\bar{8}{\cdot}7$
	12	$\overline{13}{\cdot}15$	$\overline{12}{\cdot}77$				
$t=1{\cdot}5$	2	1·45	1·17				
	4	0·24	0·16				
	6	$\bar{2}{\cdot}35$	$\bar{2}{\cdot}63$				
	8	$\bar{4}{\cdot}26$	$\bar{3}{\cdot}14$				
	10	$\bar{6}{\cdot}11$	$\bar{5}{\cdot}19$				

Digits in bold type before the decimal place † imply that the decimal is to be multiplied by this power of ten; thus $\bar{1}{\cdot}45$ means 0·045 and 1·45 means 4·5.

† The notation is due to A. C. Aitken (1925, footnote to p. 293).

for several values of t. These are given for s even, since in this case $\bar{R}(t)$ always lies between r_s and r_{s-1} for $0<t<\sqrt{3}$ (and sooner or later for all $t>0$). Similar remarks apply to T_s and T_{s-1}. Thus the error in the approximations r_s and T_s is less in value than $|\nabla r_s|$ and $|\nabla T_s|$ respectively. It will be observed that the C.F. converges ultimately more rapidly than the series, as far as the present analysis is concerned. Whether this state of affairs holds for larger values of s than those considered can only be conjectured. For the early convergents the series has a slight advantage, especially for small t; but for $t>1$ the C.F. soon becomes more accurate, although for larger values of t there is the disadvantage of the negative approximations in the early stages. However, this disadvantage may be overcome by using the generalization of (1), namely,

$$\bar{R}(t)=\sum_{r=0}^{s-1}\frac{t^{2r+1}}{1.3.5.....(2r+1)}+\frac{t^{2s}}{1.3.....(2s+1)}\left[\frac{t}{2s+1}-\frac{(2s+1)\,t^2}{2s+3}+\frac{2t^2}{2s+5}-\frac{(2s+3)\,t^2}{2s+7}+\frac{4t^2}{2s+9}-\ldots\right]. \tag{11}$$

If s is taken large enough in this expression (which converges for $t>0$) negative approximations are avoided. Numerical examples support the view that the rate of convergence of (11) is similar to that of (1), but from an analytical point of view there are difficulties.

The series (2) in its rate of convergence is not a serious rival to the expansions considered in Table 2. Thus the C.F. appears to be on the whole as good an approximation to $\bar{R}(t)$ as any other.

3. The Laplace continued fraction for $R(t)$

It can be shown (see, for example, Bromwich, 1926, p. 388) that

$$R(t)=t\int_{-\infty}^{\infty}\frac{g(x)\,dx}{x^2+t^2}=\frac{t}{2\sqrt{\pi}}\int_0^{\infty}\frac{e^{-x}x^{-\frac{1}{2}}}{x+\frac{1}{2}t^2}\,dx, \tag{12}$$

so that
$$R(t)=\frac{1}{t}-\frac{1}{t^3}+\frac{1.3}{t^5}-\ldots+(-)^s\frac{1.3.5.....(2s-1)}{t^{2s+1}}+R_s(t), \tag{13}$$

where
$$|R_s(t)|=t^{-(2s+1)}\int_{-\infty}^{\infty}\frac{x^{2s+2}g(x)\,dx}{x^2+t^2}<\frac{1.3.5.....(2s+1)}{t^{2s+3}},$$

which states the familiar fact that the error involved in using any number of terms of (13) is always less in value than the first term omitted.

In terms of the hypergeometric function we have

$$R(t)\doteqdot\frac{1}{t}-\frac{1}{t^3}+\frac{1.3}{t^5}-\ldots+(-)^s\frac{1.3.....(2s-1)}{t^{2s+1}}+(-)^{s+1}\frac{1.3.....(2s+1)}{t^{2s+3}}\lim_{c\to\infty}F\left(1,s+\tfrac{3}{2};c;-\frac{2c}{t^2}\right), \tag{14}$$

so that using the C.F. of Gauss for this case we find

$$R(t)=\frac{1}{t}-\frac{1}{t^3}+\frac{1.3}{t^5}-\ldots+(-1)^s\frac{1.3.....(2s-1)}{t^{2s+1}}+(-)^{s+1}\frac{1.3.....(2s+1)}{t^{2s+2}}\left[\frac{1}{t}+\frac{2s+3}{t}+\frac{2}{t}+\frac{2s+5}{t}+\frac{4}{t}+\frac{2s+7}{t}+\ldots\right], \tag{15}$$

the C.F. part of this converging for $t > 0$, the odd convergents forming a decreasing sequence, and the even convergents an increasing sequence. If we take $s+1=0$ in (15) we derive the Laplace (1805) C.F.

$$R(t) = \frac{1}{t}+\frac{1}{t}+\frac{2}{t}+\frac{3}{t}+\frac{4}{t}+\frac{5}{t}+\ldots. \tag{16}$$

The following results are required in the sequel:

(i) Denoting the sth convergent of (16) by $\phi_s = \chi_s/\omega_s$, we see that χ_s and ω_s follow the recurrence relation

$$\left.\begin{aligned} y_s &= ty_{s-1}+(s-1)y_{s-2} \quad (s=2,3,\ldots),\\ \chi_0 &= 0, \quad \chi_1 = 1, \quad \omega_0 = 1, \quad \omega_1 = t. \end{aligned}\right\} \tag{17}$$

From (17) it is easily shown that

$$\left.\begin{aligned} \phi_s-\phi_{s-1} &= (-)^{s-1}(s-1)!/(\omega_{s-1}\omega_s),\\ \phi_s-\phi_{s-2} &= (-)^{s-1}t(s-2)!/(\omega_{s-2}\omega_s), \end{aligned}\right\} \tag{18}$$

so that we have $$\phi_0 < \phi_2 < \phi_4 \ldots < R(t) < \ldots < \phi_5 < \phi_3 < \phi_1. \tag{19}$$

(ii) The integral for $R(t)$ may be written

$$R(t) = \int_0^\infty \exp(-\tfrac{1}{2}x^2 - xt)\,dx, \tag{20}$$

which after differentiation and the use of (17) yields

$$R\omega_s - \chi_s = (-)^s \int_t^\infty (x-t)^s g(x)\,dx/g(t). \tag{21}$$

(iii) It is well known that ω_s is a Hermite polynomial and satisfies the differential equation

$$\frac{d^2\omega_s}{dt^2} + t\frac{d\omega_s}{dt} - s\omega_s = 0. \tag{22}$$

Or if we set $\omega_s = e^{-\frac{1}{4}t^2}\varpi_s$, then $$\frac{d^2\varpi_s}{dt^2} = (\tfrac{1}{4}t^2 + s + \tfrac{1}{2})\varpi_s. \tag{23}$$

We require an asymptotic formula for ϖ_s for large s. Following a method outlined in Jeffreys & Jeffreys (1946, § 17·122) we put $\varpi_s = \exp\left\{\int_0^t \eta(\theta)\,d\theta\right\}$, so that $\eta^2 + \frac{d\eta}{dt} = \frac{1}{4}t^2 + s + \frac{1}{2}$. This leads to the asymptotic expressions (in s)

$$\left.\begin{aligned} \omega_{2s} &\doteqdot e^{-\frac{1}{4}t^2}\frac{1.3.5.\ldots.(2s-1)}{[1+t^2/(8s+2)]^{\frac{1}{4}}}\cosh\{\lambda_{2s}(t)\},\\ \omega_{2s+1} &\doteqdot e^{-\frac{1}{4}t^2}\frac{1.3.5.\ldots.(2s+1)}{[1+t^2/(8s+6)]^{\frac{1}{4}}}\frac{\sinh\{\lambda_{2s+1}(t)\}}{\sqrt{(2s+\frac{3}{2})}}, \end{aligned}\right\} \tag{24}$$

where $$\lambda_s(t) = \tfrac{1}{2}t\sqrt{(\tfrac{1}{4}t^2+s+\tfrac{1}{2})} + (s+\tfrac{1}{2})\ln\{\sqrt{(1+t^2/(4s+2))} + t/\sqrt{(4s+2)}\}.$$

Using an approximation to the factorial we find the following approximations to the difference between two convergents:

$$\left.\begin{aligned} \phi_{2s+1}-\phi_{2s} &\doteqdot 2\sqrt{(2\pi)}\{(1+t^2/(8s+2))\,(1+t^2/(8s+6))\}^{\frac{1}{4}}\exp\{\tfrac{1}{2}t^2 - \tfrac{1}{2}t\sqrt{(8s+2)} - \tfrac{1}{2}t\sqrt{(8s+6)}\},\\ \phi_{2s-1}-\phi_{2s} &\doteqdot 2\sqrt{(2\pi)}\{(1+t^2/(8s-2))\,(1+t^2/(8s+2))\}^{\frac{1}{4}}\exp\{\tfrac{1}{2}t^2 - \tfrac{1}{2}t\sqrt{(8s-2)} - \tfrac{1}{2}t\sqrt{(8s+2)}\}. \end{aligned}\right\} \tag{25}$$

The more complicated formulae for these differences found by using (24) in (18) is perhaps more reliable when t has a moderate value such as greater than three.

4. Numerical comparison of three expansions for $R(t)$

We shall consider the expansions

$$R(t) = \frac{1}{t} + \frac{1}{t} + \frac{2}{t} + \frac{3}{t} + \ldots, \tag{A}$$

$$\bar{R}(t) = \frac{t}{1} - \frac{t^2}{3} + \frac{2t^2}{5} - \frac{3t^2}{7} + \ldots, \tag{B}$$

$$\bar{R}(t) = e^{\frac{1}{2}t^2}\left[\frac{t}{1} - \frac{t^3}{3.2.1!} + \frac{t^5}{5.2^2.2!} - \frac{t^7}{7.2^3.3!} + \ldots\right], \tag{C}$$

each of which can be used to approximate to $R(t)$ through the relation $R(t) + \bar{R}(t) = 1/(2g)$, from the point of view of rapidity of convergence. We consider how many terms are required so that the error is $10^{-(n+1)} \times 2{\cdot}5$. More precisely, if C_{s-1} and C_s are successive convergents of a C.F. (or partial sums of a series) providing lower and upper bounds, then we evaluate the least value of s (an integer) for which $|C_{s-1} - C_s| \leqslant 10^{-(n+1)} \times 2{\cdot}5$ is true approximately. This is about equivalent to ensuring accuracy in the nth place of decimals. For the Laplace C.F. we use the approximation

$$2\sqrt{(2\pi)}\exp\{\tfrac{1}{2}t^2 - \tfrac{1}{2}t\sqrt{(4s+2)} - \tfrac{1}{2}t\sqrt{(4s-2)}\} = 10^{-(n+1)} \times 2{\cdot}5,$$

the solution of which is given by

$$s = \psi(t) + 1/[16\psi(t)], \quad \text{where} \quad \psi(t) = [\tfrac{1}{4}t + (1{\cdot}4992 + 1{\cdot}1513n)/t]^2,$$

where the second term is unimportant for small t. As a simple crude formula for (A) we have $|\phi_{s-1} - \phi_s| = 2\sqrt{(2\pi)}\exp(\frac{1}{2}t^2 - 2t\sqrt{s})$. Thus, approximately, the necessary number of terms varies directly as the square of the accuracy, and inversely as the square of t. We have already seen from § 2 that the rates of convergence of (B) and (C) are very similar. Indeed, for small values of t we have $r_{4s+2} - r_{4s+1} = \pi(T_{4s+2} - T_{4s+1})/2^{4s-1}$, so that (B) has an advantage in view of the factor 2^{4s-1}. A similar relation for t not small appears to be complicated.

A comparison of the three expansions is given in Table 3.

Table 3. *Number of terms to achieve a given accuracy for three expansions*

	Accuracy					
	$2{\cdot}5 \times 10^{-7}$			$2{\cdot}5 \times 10^{-13}$		
Expansion	A	B	C	A	B	C
t						
0·1	7,072	4	3	23,460	6	5
0·25	1,135	4	4	3,760	8	7
0·5	287	6	5	946	10	9
0·75	130	6	7	425	10	11
1·0	75	8	8	242	12	13
1·5	35	10	11	112	14	16
2·0	22	14	15	66	18	21
2·5	15	16	19	45	22	26
3·0	11	20	24	30	26	31
4·0	9	—	37	21	—	45
5·0	7	—	53	17	—	62
10·0	4	—	187	9	—	197

Remarks on Table 3

First of all it must be remarked that precise accuracy in a table of this kind is unnecessary. Most of the numbers in the region of 25 or less were found by direct calculation, but the rest are approximate. Again there is a slight difference between finding the number of terms (s) of an expansion to ensure an error less than ϵ when the true value is known, as compared to relying on the expansion alone to give this information. From this point of view since the terms in (B) are in *pairs* greater than and less than the true value, the numbers given often give greater accuracy than is quoted. Thus four terms of (B) give accuracy in the ninth decimal place.

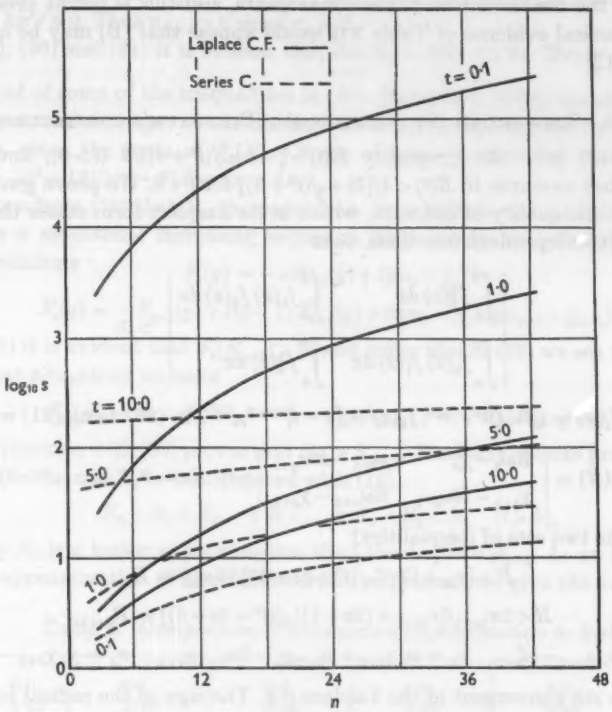


Fig. 1. Number of terms (s) for accuracy to be of the order $10^{-(n+1)} \times 2{\cdot}5$ for the Laplace C.F. and for series C.

It will be seen that the number of terms for a required accuracy increases with decreasing t for the Laplace C.F., the increase being rapid as t approaches zero, while for (B) and (C) there is a steady increase with increasing t. Again for (A) the ratio of the number of terms for 12 places of decimals as compared to 6 is approximately three, whereas the corresponding factor for (B) and (C) is less and decreases as t increases. Expressed otherwise, we may say that increasing accuracy for the Laplace C.F. for a given value of t is only achieved at the expense of an increasing additional number of terms. To take an extreme case, we remark that for $t = 0{\cdot}1$, the additional numbers of terms required for successive jumps of six places

of decimals, starting from zero, are: 7072, 16390, 25934, 35477, 45020, 54564 approximately. For the series (C) the corresponding numbers for the case $t = 10$ are: 187, 10, 10, 10, 9, 9.

A graphical comparison of the expansions (A) and (C) showing the logarithm of the number of terms against the accuracy is given in Fig. 1.† It is evident from this that the usefulness of the series (C) falls off slowly with increasing t, and from the flatness of the curves when n is large (for a fixed t) it is clear that although convergence may be slow at first, sooner or later it becomes rapid. By contrast the Laplace C.F. falls off rapidly in its usefulness as t decreases, and there is little tendency to flatness for increasing n when t is fixed.

To discuss the behaviour of (B) for moderate or large values of t requires an asymptotic expression for the denominators of the convergents, and this is not at present available. From the numerical evidence of Table 3 it would appear that (B) may be quicker in converging than (C).

5. Schwarzian inequalities and Birnbaum's conjecture

Birnbaum (1942) gave the inequality $R(t) > [-t+\sqrt{(t^2+4)}]/2$ $(t>0)$, and later (1950) conjectured what amounts to $R(t) < 4/[3t+\sqrt{(t^2+8)}]$ for $t>0$. We prove generalizations of these, using the inequality of Schwarz, which in its simplest form states that if $f_1(x)$ and $f_2(x)$ are linearly independent functions, then

$$\begin{vmatrix} \int_a^b f_1^2(x)\,dx & \int_a^b f_1(x)\,f_2(x)\,dx \\ \int_a^b f_2(x)\,f_1(x)\,dx & \int_a^b f_2^2(x)\,dx \end{vmatrix} > 0. \tag{26}$$

Hence taking $f_1(x) = (x-t)^{\frac{1}{2}s}\,e^{-\frac{1}{4}x^2}$, $f_2(x) = (x-t)^{\frac{1}{2}(s+2)}\,e^{-\frac{1}{4}x^2}$ in (26) using (21) we find

$$F_s(R) = \begin{vmatrix} R\omega_s - \chi_s & \chi_{s+1} - R\omega_{s+1} \\ \chi_{s+1} - R\omega_{s+1} & R\omega_{s+2} - \chi_{s+2} \end{vmatrix} > 0 \quad (s = 0, 1, 2, \ldots;\ t \geqslant 0). \tag{27}$$

This leads to the two sets of inequalities‡

$$R > [\nu_{2s} + (2s)!\,\sqrt{(t^2+8s+4)}]/(2\mu_{2s}) = R_{2s}, \tag{28a}$$

$$R < 2\pi_{2s+1}/[\nu_{2s+1} + (2s+1)!\,\sqrt{(t^2+8s+8)}] = R_{2s+1}, \tag{28b}$$

where $\mu_s = \omega_s\omega_{s+2} - \omega_{s+1}^2$, $\nu_s = \omega_s\chi_{s+2} + \omega_{s+2}\chi_s - 2\omega_{s+1}\chi_{s+1}$, $\pi_s = \chi_s\chi_{s+2} - \chi_{s+1}^2$,

and χ_s/ω_s is the sth convergent of the Laplace C.F. The sign of the radical in (28) is easily determined provided we can show that for $t \geqslant 0$, (i) $\mu_{2s} > 0$ and (ii) $\pi_{2s+1} > 0$. For if this is the case we see that since

$$F_s(\phi_s) < 0, \quad F_s(\phi_{s+2}) < 0, \quad F_s(\phi_{s+1}) > 0, \tag{29}$$

and also from (19) that R lies between ϕ_{s+1} and ϕ_{s+2}, then

$$\phi_{2s+2} < R_{2s} < R < \phi_{2s+1}, \tag{30}$$

† I am indebted to Dr J. A. Storrow for drawing this diagram.

‡ When $t = 0$, $R_{2s}(0) = \dfrac{2^{2s-1}s!\,(s-1)!}{(2s-1)!\,\sqrt{(2s+1)}}$, so that in the limit $R(0) = \sqrt{(\frac{1}{2}\pi)}$ (from the formula of Wallis), providing a new demonstration of the value of the total area under the normal curve. A similar result follows from $R_{2s+1}(0)$.

where R_{2s} is the larger root of $F_{2s}(R) = 0$. Similarly, if (ii) holds, then by considering the sign of $\hat{R}^2 F_{2s+1}(\hat{R}^{-1})$ when $\hat{R}^{-1} = \phi_{2s+1}, \phi_{2s+2}, \phi_{2s+3}$, we see that

$$\phi_{2s+2} < R < R_{2s+1} < \phi_{2s+3}, \tag{31}$$

where R_{2s+1} is the smaller root of $F_{2s+1}(R) = 0$.

To prove (i) we have the integral representation $\omega_s = \int_{-\infty}^{\infty} g(x)\,(t-x)^s\,dx$, so that the result follows immediately from the inequality of Schwarz. For (ii) there is the recurrence relation

$$\pi_s = (s^2-1)\,\pi_{s-2} + \chi_{s-1}(\chi_{s+1} - \overline{s+1}\,\chi_{s-1}),$$

and if $\sigma_s = \chi_s - s\chi_{s-2}$ then $\sigma_{s+2} = (s-1)\,\sigma_s + t^2\chi_s$. But $\sigma_2 = t$, $\sigma_4 = t^3 + t$, and so $\sigma_{2s} \geqslant 0$ since $\chi_{2s} \geqslant 0$ for $t \geqslant 0$. Thus $\pi_{2s+1} > 0$ since $\pi_1 = 2$.

From (19), (30) and (31) it is evident that $\lim_{s\to\infty} R_s = R(t)$ $(t \geqslant 0)$. Murty (1952) has given a partial proof of some of the inequalities in (28). Sampford (1953) has shown that $R < R_1$ holds for the extended range $t > -1$. Similar extensions could be given for other cases also; for example, since the roots of $F_2(R) = 0$ are of opposite sign for $0 > t > -2$, we have $R > [t^3 - 2t + \sqrt{(t^2+12)}]/(t^4+3)$ for $t > -2$.

It is obvious from (30) that R_{2s} increases from time to time with s; but it is not obvious that $\{R_{2s}\}$ is a monotonic increasing sequence. This can be proved by considering the recurrence relations

$$F_s(y) = -sF_{s-1}(y) + (y\omega_s - \chi_s)^2 \tag{32}$$

and

$$F_s(y) = -F_{s-1}(y) + s(s-1)\,F_{s-2}(y) + t(y\omega_s - \chi_s)\,(y\omega_{s-1} - \chi_{s-1}). \tag{33}$$

For from (32) it is evident that $F_s(R_{s-1}) > 0$, and using this in (33) we see that $F_s(R_{s-2}) < 0$. It follows that altogether we have

$$F_{2s}(\phi_{2s}) < 0, \quad F_{2s}(R_{2s}) = 0, \quad F_{2s}(\phi_{2s+1}) > 0, \quad F_{2s}(R_{2s-2}) < 0,$$

which in conjunction with (30) proves that $R_{2s} > R_{2s-2}$. Similarly, we can prove $R_{2s+3} < R_{2s+1}$. Hence we have the result, to be compared with (19),

$$R_0 < R_2 < R_4 \ldots < R < \ldots < R_5 < R_3 < R_1 \quad (t \geqslant 0), \tag{34}$$

in which any R_s is a better approximation than the Laplace ϕ_{s+2}. As an indication of the closeness of approximation of these Schwarzian sequences, we give the value of the error

Table 4. *Comparison of Schwarzian approximation to $R(t)$*

t ... / R ...	0·0 / 1·25	0·1 / 1·16	0·5 / 0·88	1·0 / 0·66	2·0 / 0·42	3·0 / 0·30
R_0	0·25	0·21 (1·11)	$\bar{1}$·96 (0·48)	$\bar{1}$·38 (0·15)	$\bar{2}$·72 ($\bar{1}$·21)	$\bar{2}$·18 ($\bar{2}$·46)
R_1	0·16	0·12 (1·55)	$\bar{1}$·38 (0·50)	$\bar{1}$·11 (0·10)	$\bar{2}$·13 ($\bar{2}$·74)	$\bar{3}$·22 ($\bar{2}$·10)
R_2	$\bar{1}$·99	$\bar{1}$·70 (1·10)	$\bar{1}$·19 (0·30)	$\bar{2}$·42 ($\bar{1}$·56)	$\bar{3}$·32 ($\bar{2}$·28)	$\bar{4}$·36 ($\bar{3}$·24)
R_3	$\bar{1}$·80	$\bar{1}$·53 (1·42)	$\bar{1}$·11 (0·29)	$\bar{2}$·20 ($\bar{1}$·37)	$\bar{4}$·95 ($\bar{2}$·12)	$\bar{5}$·74 ($\bar{4}$·69)
R_4	$\bar{1}$·61	$\bar{1}$·39 (0·95)	$\bar{2}$·73 (0·20)	$\bar{2}$·10 ($\bar{1}$·24)	$\bar{4}$·53 ($\bar{3}$·53)	$\bar{5}$·18 ($\bar{4}$·22)
R_5	$\bar{1}$·53	$\bar{1}$·32 (1·34)	$\bar{2}$·49 (0·19)	$\bar{3}$·57 ($\bar{1}$·17)	$\bar{4}$·13 ($\bar{3}$·25)	$\bar{6}$·49 ($\bar{5}$·74)
P.W.	0·00	$\bar{5}$·76	$\bar{3}$·93	$\bar{2}$·73	$\bar{1}$·51	0·12

Figures in the body of the table represent |Approximation − True value| with the same convention as in Table 2. Figures in parentheses in the same line as R_s, $(s = 0, 1, \ldots, 5)$, refer to ϕ_{s+2}. P.W. refers to the Pólya-Williams inequality.

of R_s, $s = 0$ to 5, for six values of t in Table 4. For comparison we also give the error for the Laplace convergent ϕ_{s+2}, and that for the Pólya (1945) and Williams (1946) inequality

$$\int_0^t g(x)\,dx < \tfrac{1}{2}[1-\exp(-2t^2/\pi)]^{\frac{1}{2}} \quad (t \geqslant 0).$$

Strictly speaking the latter is a different type of inequality, and although Pólya gives a generalization of it, convergency properties have not been considered.

Remarks on Table 4

The second row gives the value of $R(t)$ to two decimal places merely to indicate the importance of the errors in the approximations. It will be observed that the Schwarzian values are always better than the corresponding Laplace convergents, and this is particularly so for small values of t; thus for $t = 0{\cdot}1$ the error in R_4 is within about 4 %, whereas the error in ϕ_6 is in the region of 100 %. It is also interesting to notice that the P.W. values are better than the other approximations for $t < 1{\cdot}0$, but deteriorate for $t > 1{\cdot}0$. Of course for $t < 1{\cdot}0$ higher Schwarzian approximations would ultimately become better than the P.W. inequality.

6. Summary and conclusion

We have given a new continued fraction (C.F.) for the normal integral ratio

$$\bar{R}(t) = e^{\frac{1}{2}t^2}\int_0^t e^{-\frac{1}{2}x^2}\,dx,$$

which turns out to be rapidly convergent for small values of t (say less than $\sqrt{3}$); for moderate values convergence is slow at first but becomes quite rapid in due course; at least this is supported by numerical evidence. The rate of convergence of this C.F. is compared with a series and also with the Laplace C.F. for the Mills's ratio $R(t) = e^{\frac{1}{2}t^2}\int_t^\infty e^{-\frac{1}{2}x^2}\,dx$. An interesting point about the Laplace C.F. is that the difference between the sth and $(s-1)$th convergents has the approximate magnitude $2\sqrt{(2\pi)}\exp(\frac{1}{2}t^2 - 2t\sqrt{s})$, when t is not large. This, together with numerical evidence, shows that fair accuracy is attainable with a few terms when t is moderate to large, but the rate of convergence deteriorates rapidly as t becomes small. In addition, for a fixed value of t (not large) the necessary number of terms for a given accuracy varies directly as the *square of the accuracy*. The behaviour of the new C.F. for $\bar{R}(t)$ is completely different, for it converges rapidly for small t and deteriorates slowly as t increases, whereas for a fixed value of t, increasing accuracy seems to demand much less in terms than the sequence of the accuracy.

We also give two sets of inequalities for $R(t)$, consisting of irrational fractions, these being generalizations of results by Birnbaum. We prove that one set increases, and the other decreases, monotonically, to the limit $R(t)$. A particular case includes the formula for the total area under the normal curve.

In an appendix we give some properties of the C.F.'s for $\bar{R}(t)$ and $R(t)$ which are not readily available in the literature, and a brief summary of expansions and inequalities.

APPENDIX

(1) *Properties of the Laplace continued fraction for $R(t)$*

(*a*) (i) $R(t) = \sum_{s=0}^{\infty} (-)^s s!/(\omega_s \omega_{s+1})$.

(ii) $R(t) = t \sum_{s=0}^{\infty} (2s)!/(\omega_{2s}\omega_{2s+2})$.

(iii) $R(t) = 1/\omega_1 - t \sum_{s=0}^{\infty} (2s+1)!/(\omega_{2s+1}\omega_{2s+3})$.

These follow from the general theory of C.F.'s.

(*b*) (i) $\chi_{s+1} = \omega_s + \chi'_s$.

(ii) $\chi_{s+1} = \omega_s + \omega'_{s-1} + \omega''_{s-2} + \ldots$.

Inductional methods may be used for (i), from which (ii) is deduced. A general theorem, of which (ii) is a particular case, using more elaborate methods is given by Sherman (1933).

(*c*) $$\frac{R\omega_s - \chi_s}{\chi_{s-1} - R\omega_{s-1}} = \frac{s}{t} + \frac{s+1}{t} + \frac{s+2}{t} + \ldots \quad (t>0).$$

The rth convergent of the C.F. leads to the approximation ϕ_{r+s} to $R(t)$. This result may be used as a computational check.

(*d*) Remainder formulae:

$$R\omega_{2s} - \chi_{2s} = \frac{2^{s-1}s!\,t}{\sqrt{\pi}} \int_0^\infty \frac{e^{-x} x^{s-\frac{1}{2}}}{(x+\frac{1}{2}t^2)^{s+1}}\,dx \quad (s = 0, 1, 2, \ldots),$$

$$\chi_{2s-1} - R\omega_{2s-1} = \frac{2^{s-1}(s-1)!}{\sqrt{\pi}} \int_0^\infty \frac{e^{-x} x^{s-\frac{1}{2}}}{(x+\frac{1}{2}t^2)^{s}}\,dx \quad (s = 1, 2, \ldots; t>0).$$

These follow from the general theory of C.F.'s.

(*e*) Systematic summation of asymptotic series for $R(t)$:

$$\frac{1}{t} - \frac{1}{t^3} + \frac{1.3}{t^5} - \ldots + (-1)^s \frac{1.3\ldots.(2s-3)}{t^{2s-1}} = \frac{t}{t^2} + \frac{t^2}{t^2-1} + \frac{3t^2}{t^2-3} + \ldots + \frac{(2s-3)t^2}{t^2-(2s-1)}.$$

(2) *Properties of the new continued fraction for $\bar{R}(t)$*

$$\bar{R}(t) = \frac{t}{1} - \frac{t^2}{3} + \frac{2t^2}{5} - \frac{3t^2}{7} + \ldots = \lim_{s\to\infty} r_s = \lim_{s\to\infty} p_s/q_s \quad (t>0).$$

(*a*) Differences of convergence:

(i) $r_{2s+1} - r_{2s} = (-2)^s s!(1.3\ldots.(2s-1))\,t^{4s+1}/(q_{2s}q_{2s+1})$.

(ii) $r_{2s} - r_{2s-1} = (-2)^{s-1}(s-1)!\,(1.3\ldots.(2s-1))\,t^{4s-1}/(q_{2s-1}q_{2s})$.

(iii) $r_{2s} - r_{2s-2} = (-)^{s-1}(4s-1)\,(2s-2)!\,t^{4s-3}/(q_{2s-2}q_{2s})$.

(iv) $r_{2s+1} - r_{2s-1} = (-)^{s-1}(4s+1)\,(2s-1)!\,t^{4s-1}/(q_{2s-1}q_{2s+1})$.

(*b*) Formulae for denominators:

(i) $q_{2s} = \int_{-\infty}^{\infty} g(x)\,x^{2s}(x^2-t^2)^s\,dx$.

(ii) $q_{2s+1} = \int_{-\infty}^{\infty} g(x)\,x^{2s+2}(x^2-t^2)^s\,dx$.

(*c*) Remainder formulae:

$$\text{(i)}\quad \bar{R}q_{2s}-p_{2s}=\int_0^t x^{2s}(x^2-t^2)^s\exp\tfrac{1}{2}(t^2-x^2)\,dx.$$

$$\text{(ii)}\quad \bar{R}q_{2s+1}-p_{2s+1}=\int_0^t x^{2s+2}(x^2-t^2)^s\exp\tfrac{1}{2}(t^2-x^2)\,dx.$$

(*d*) Formulae for computational checks:

$$\text{(i)}\quad \frac{\bar{R}q_{2s+1}-p_{2s+1}}{\bar{R}q_{2s}-p_{2s}}=\frac{(2s+1)\,t^2}{4s+3}-\frac{(2s+2)\,t^2}{4s+5}+\frac{(2s+3)\,t^2}{4s+7}-\ldots$$

n terms leading to the approximation r_{2s+n+1} for $\bar{R}(t)$.

$$\text{(ii)}\quad \frac{\bar{R}q_{2s}-p_{2s}}{p_{2s-1}-\bar{R}q_{2s-1}}=\frac{2st^2}{4s+1-}\frac{(2s+1)\,t^2}{4s+3}+\frac{(2s+2)\,t^2}{4s+5}-\ldots$$

n terms leading to the approximation r_{2s+n} for $\bar{R}(t)$.

(*e*) Partial sums of series for $\bar{R}(t)$:

$$\sum_{r=0}^{s-1}\frac{t^{2r+1}}{1.3.\ldots.(2r+1)}=\frac{t}{1-}\frac{t^2}{t^2+3-}\frac{3t^2}{t^2+5-}\ldots-\frac{(2s-3)\,t^2}{t^2+2s-1}.$$

(3) *Miscellaneous results*

(*a*) Factorial type of series for $R(t)$ due to Schlömilch (1895)

$$tR(t)=1+\frac{1}{\sqrt{\pi}}\sum_{r=1}^{\infty}(-2)^r\int_0^{\infty}x^{(r)}e^{-x}x^{-\frac{1}{2}}dx/(u_1u_2\ldots u_r),$$

where $$u_r=t^2+2r,\quad x^{(r)}=x(x-1)\ldots(x-r+1)\quad(t>0).$$

(This result is mentioned by Wishart (1927).)

In particular $$tR(t)=1-\frac{1}{u_1}+\frac{1}{u_1u_2}-\frac{5}{u_1u_2u_3}+\frac{9}{u_1u_2u_3u_4}-\frac{129}{u_1u_2u_3u_4u_5}+\ldots.$$

(*b*) $$\int_0^t g(x)\,dx<a_n^{-1}(2\pi)^{-\frac{1}{2}}\left[n\int_0^{ta_n}x^{n-1}e^{-\frac{1}{2}x^2}dx\right]^{1/n}\quad(t>0),$$

where $$a_n=2\pi^{-\frac{1}{2}}\left[\Gamma\left(\frac{n}{2}+1\right)\right]^{1/n}\quad(n=2,3,\ldots).$$

Due to Pólya (1945), the particular case $n=2$ being given by Williams (1946). (See also Tate (1953).)

(*c*) Minimum property of the convergents of the Laplace C.F.:

$$\min_{\pi_s}\frac{t}{2\sqrt{\pi}}\int_0^{\infty}e^{-x}x^{-\frac{1}{2}}(x+\tfrac{1}{2}t^2)\left\{\frac{1}{x+\frac{1}{2}t^2}-\pi_s(x)\right\}^2dx=R(t)-\chi_{2s+2}/\omega_{2s+2},$$

$$\min_{\pi_s}\frac{1}{\sqrt{\pi}}\int_0^{\infty}e^{-x}x^{\frac{1}{2}}(x+\tfrac{1}{2}t^2)\left\{\frac{1}{x+\frac{1}{2}t^2}-\pi_s(x)\right\}^2dx=t[\chi_{2s+2}/\omega_{2s+2}-R(t)],$$

the minimum being taken over all polynomials of precise degree s. These are implied in a general theorem due to Stieltjes. (See, for example, Shohat & Tamarkin (1943), p. 75.) In a similar way we find the following inequalities:

(i) $(t^2+1)\,R(t)-t>2t/(t^4+6t^2+15)$,

(ii) $(t^2+1)\,R(t)-t>2(t^5+14t^3+75t)/(t^6+20t^4+150t^2+420t^2+525)$,

(iii) $(t^2+3t)\,R(t)<t^2+2-3!/(t^4+10t^2+35)$,

(iv) $(t^3+3t)\,R(t)<t^2+2-3!(t^4+18t^2+119)/(t^6+28t^4+294t^2+1260t^2+2205)$,

(v) $(t^4+6t^2+3)\,R(t)-t^3-5t>4!\,t/(t^6+15t^4+105t^2+315)$,

(vi) $(t^5+10t^3+15t)\,R(t)<t^4+9t^2+8-5!/(t^6+21t^4+189t^2+693)$,

where $t>0$ throughout.

(*d*) The weight function in probit analysis.

If $$\psi(t) = e^{-t^2} \Big/ \int_{-\infty}^{t} e^{-\frac{1}{2}x^2}\,dx \int_{t}^{\infty} e^{-\frac{1}{2}x^2}\,dx,$$

then $\psi(t)$ is a decreasing function of t^2 (Hammersley, 1950; Tate, 1953).

(*e*) Rough formula for the normal distribution function:

$$-10\log\int_t^\infty g(x)\,dx = 2\tfrac{1}{3}t^2 + 4 + 10\log t \quad \text{with an error less than 1 if } 2 \leqslant t \leqslant 14$$

(Good, 1950).

REFERENCES

AITKEN, A. C. (1925). *Proc. Roy. Soc. Edinb.* **46**, 289.

BIRNBAUM, Z. W. (1942). *Ann. Math. Statist.* **13**, 245.

BIRNBAUM, Z. W. (1950). *Ann. Math. Statist.* **21**, 272.

BROMWICH, T. J. I'A. (1926). *Theory of Infinite Series.* Macmillan.

GAUSS, C. F. (1812). *Werke*, **3**. Gottingen.

GOOD, I. J. (1950). *Probability and the Weighing of Evidence.* London: Griffin.

HAMMERSLEY, J. M. (1950). *J. R. Statist. Soc.* B, **12**, 209.

JEFFREYS, H. & JEFFREYS, B. S. (1946). *Methods of Mathematical Physics.* Cambridge University Press.

LANE, R. E. (1944). *Frat. Actuarial Ass., Proc.* no. **19**, p. 52.

LAPLACE, M. (1805). *Traité de Mécanique Céleste*, **4**. Paris.

MURTY, V. N. (1952). *J. Indian Soc. Agric. Statist.* **4**, no. 1, p. 84.

PERRON, O. (1913). *Die Lehre von den Kettenbruchen.* Berlin: Teubner.

PÓLYA, G. (1945). *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press, 1949.

SAMPFORD, M. R. (1953). *Ann. Math. Statist.* **24**, 130.

SCHLÖMILCH, OSKAR (1895). *Compendium der Höheren Analysis*, **2**, 265–70.

SHERMAN, J. (1933). *Trans Amer. Math. Soc.* **35**, 64–87.

SHOHAT, J. A. & TAMARKIN, J. D. (1943). *The Problem of Moments.* Mathematical Surveys, **1**, American Mathematical Society.

TATE, R. F. (1953). *Ann. Math. Statist.* **24**, 132.

TEICHROEW, D. (1952). *Mathematical Tables and other Aids to Computation*, **6**, no. 39, p. 127.

WALL, H. S. (1948). *Analytic Theory of Continued Fractions.* New York: Van Nostrand.

WILLIAMS, J. D. (1946). *Ann. Math. Statist.* **17**, 363.

WISHART, J. (1927). *Biometrika*, **19**, 13.

A CONFIDENCE REGION FOR THE SOLUTION OF A SET OF SIMULTANEOUS EQUATIONS WITH AN APPLICATION TO EXPERIMENTAL DESIGN*

By G. E. P. BOX, *Imperial Chemical Industries, Blackley, Manchester, England, and The Institute of Statistics, Raleigh, N.C.*

and J. S. HUNTER, *The Institute of Statistics, Raleigh, N.C.*

1. Introduction

The problem of finding limits of error for the solutions of a set of k linear equations obtained by equating each of the quantities

$$\left.\begin{aligned} a_{10}+a_{11}x_1+a_{12}x_2+\ldots+a_{1k}x_k &= \delta_1 \\ a_{20}+a_{21}x_1+a_{22}x_2+\ldots+a_{2k}x_k &= \delta_2 \\ \vdots \quad\quad \vdots \quad\quad\quad\quad & \quad\vdots \\ a_{k0}+a_{k1}x_1+a_{k2}x_2+\ldots+a_{kk}x_k &= \delta_k \end{aligned}\right\} \tag{1}$$

to zero, where $a_{10}, a_{11}, \ldots, a_{1k}, \ldots, a_{kk}$ are subject to error, was considered by Lonseth (1942) who gives a number of references to earlier writers. He obtained a series for the error of any unknown, considered as a function of the $k(k+1)$ errors in the coefficients. He also developed a criterion (which depended on the conditioning of the equations and the smallness of the errors relative to their coefficients) for the convergence of his series. Lonseth and more recent writers such as Hotelling (1943) and Turing (1948) have been chiefly concerned with the effect of rounding errors. Not infrequently, however, we are faced with the problem where large observational errors occur in the coefficients and the equations may not be well conditioned. It then seems essential to consider the errors in the solution not individually but jointly, in fact to determine a confidence region for the possible solution in the space of $x_1, x_2, \ldots, x_k$. One example of this circumstance which has attracted the attention of the authors to the problem occurs when attempts are made to attach limits of error to the position of a maximum in experiments of the type discussed by Box & Wilson (1951). Further reference will be made to this problem later.

2. An exact confidence region

We assume that the errors in the coefficients $a_{10}, a_{11}, \ldots, a_{kk}$ are distributed multinormally with a $k(k+1)$ by $k(k+1)$ variance-covariance matrix $\Omega\sigma^2$, known apart from the factor σ^2, and that an estimate s^2 of σ^2 is available based on ϕ degrees of freedom and distributed as $\chi^2(\phi)/\phi$ independently of the errors in the coefficients. We use the notation $\chi^2(\phi)$ to denote a quantity distributed in the χ^2 distribution with ϕ degrees of freedom; similarly, $F(\phi_1, \phi_2)$ denotes a quantity distributed in the Fisher-Snedecor F distribution with ϕ_1 and ϕ_2 degrees of freedom. A confidence region may now be found by an extension of an argument given by Fieller (1940).

Consider the expressions $\delta_1, \delta_2, \ldots, \delta_k$ of equation (1) for a fixed set of values $x_1^0, x_2^0, \ldots, x_k^0$. Knowing Ω we may readily calculate the variance-covariance matrix $\mathscr{E}(\boldsymbol{\delta}\boldsymbol{\delta}') = \mathbf{V}\sigma^2$, where

* Work sponsored by The Office of Ordnance Research, U.S. Army.

$\boldsymbol{\delta}$ denotes the $k \times 1$ column vector of the δ's. Since $\dfrac{\boldsymbol{\delta}'\mathbf{V}^{-1}\boldsymbol{\delta}}{ks^2}$ is distributed as $F(k, \phi)$, the probability is $1-\alpha$ that the inequality

$$\frac{\boldsymbol{\delta}'\mathbf{V}^{-1}\boldsymbol{\delta}}{ks^2} < F_\alpha(k, \phi) \tag{2}$$

is true, where F_α denotes the α probability point of the F distribution.

Now consider the space of the $k(k+2)+1$ quantities $a_{10}, a_{11}, \ldots, a_{kk}$, $s^2, x_1, x_2, \ldots, x_k$. The expression

$$\boldsymbol{\delta}'\mathbf{V}^{-1}\boldsymbol{\delta}/ks^2$$

is a function of these quantities, and equation (2) defines a region $R(x_1^0, x_2^0, \ldots, x_k^0)$ on the hyperplane $x_1 = x_1^0, x_2 = x_2^0, \ldots, x_k = x_k^0$ such that the probability that the point $a_{10}, a_{11}, \ldots, a_{kk}$, $s^2, x_1^0, x_2^0, \ldots, x_k^0$ falls within this region is $1-\alpha$. Now if we combine such regions for all values of $x_1, x_2, \ldots, x_k$ we obtain a region R in the whole $k(k+2)+1$-dimensional space which is such that the chance that a point $a_{10}, a_{11}, \ldots, a_{kk}, s^2, x_1, x_2, \ldots, x_k$ falls within R is $1-\alpha$. It follows that, given a set of observed values $a_{10}^0, a_{11}^0, \ldots, a_{kk}^0, (s^0)^2$, the region within R on the hyperplane $a_{10} = a_{10}^0$, $a_{11} = a_{11}^0, \ldots, a_{kk} = a_{kk}^0$, $s^2 = (s^0)^2$ supplies the $1-\alpha$ confidence region in the usual sense for $x_1, x_2, \ldots, x_k$. A valuable discussion of the confidence argument pertinent to this problem has been given by Bose & Roy (1953).

The limits of the $1-\alpha$ confidence region are given therefore by

$$\boldsymbol{\delta}'\mathbf{V}^{-1}\boldsymbol{\delta} = s^2 k F_\alpha(k, \phi). \tag{3}$$

Now the original quadratic form may be written as a ratio of two determinants

$$\boldsymbol{\delta}'\mathbf{V}^{-1}\boldsymbol{\delta} = \frac{-\begin{vmatrix} 0 & \boldsymbol{\delta}' \\ \boldsymbol{\delta} & \mathbf{V} \end{vmatrix}}{|\mathbf{V}|}. \tag{4}$$

Consequently, in general the boundary of the confidence region is defined by those values of $x_1, x_2, \ldots, x_k$ which cause the equation

$$\begin{vmatrix} s^2 k F_\alpha & \boldsymbol{\delta}' \\ \boldsymbol{\delta} & \mathbf{V} \end{vmatrix} = 0 \tag{5}$$

to be satisfied. If σ^2 is known, $s^2 k F_\alpha$ will be replaced by $\chi_\alpha^2 \sigma^2$ in (5).

In the important special case in which the estimates $a_{10}, a_{11}, \ldots, a_{kk}$ are uncorrelated (3) gives for the boundary of the confidence region

$$\sum_{i=1}^{k} \{(\delta_i^2)/v(\delta_i)\} = \sum_{i=1}^{k} \left\{ \left(\sum_{j=0}^{k} a_{ij} x_j \right)^2 \Big/ \sum_{j=0}^{k} v(a_{ij})\, x_j^2 \right\} = s^2 k F_\alpha, \tag{6}$$

where $v(\delta_i)\,\sigma^2$ is the variance of δ_i, $v(a_{ij})\,\sigma^2$ is the variance of a_{ij} and $x_0 \equiv 1$. If, in addition, the variances for all the coefficients are equal ($v(a_{ij}) = v(a)$; $i, j = 1, 2, \ldots, k$), the boundary of the confidence region is a quadric surface given by the equation

$$\sum \delta_i^2 = \sum_{i=1}^{k} \left(\sum_{j=0}^{k} a_{ij} x_j \right)^2 = \sum_{j=0}^{k} x_j^2\, v(a)\, s^2 k F_\alpha(k, \phi). \tag{7}$$

3. Conditioning of the equations

A circumstance which profoundly affects the nature of the confidence region is whether the equations are well or poorly 'conditioned'. We first explain what we mean by this term.

In an obvious matrix notation (1) becomes

$$\mathbf{a}_0 + \mathbf{A}\mathbf{x} = \boldsymbol{\delta}. \tag{8}$$

Consider the quadric defined by

$$\boldsymbol{\delta}'\boldsymbol{\delta} = (\mathbf{A}\mathbf{x} + \mathbf{a}_0)'(\mathbf{A}\mathbf{x} + \mathbf{a}_0). \tag{9}$$

For any fixed value of z

$$\boldsymbol{\delta}'\boldsymbol{\delta} = \sum_{i=1}^{k} \delta_i^2 = z \tag{10}$$

defines a surface which we call the conditioning surface in the $k+1$-dimensional space of $z, x_1, x_2, \ldots, x_k$. If $|\mathbf{A}| \neq 0$, in the k space of $x_1, x_2, \ldots, x_k$ the contours of (10) for fixed values of z are ellipsoids. Referring these contours to their centre $\mathbf{x}_0$ as the origin and using their principal axes as new co-ordinates $X_1, X_2, \ldots, X_k$, equation (10) may be written

$$\sum_{i=1}^{k} \delta_i^2 = \sum_{i=1}^{k} \lambda_i X_i^2 = z, \tag{11}$$

where $X_i = \mathbf{u}_i'(\mathbf{x} - \mathbf{x}_0), \mathbf{u}_i\,(i = 1, 2, \ldots, k)$ is a latent vector of $\mathbf{A}'\mathbf{A}$, and λ_i is the corresponding latent root (essentially positive). Clearly, if one or more of the latent roots are small compared with the remainder, the conditioning surface is attenuated in the direction of these axes. Thus, suppose $\lambda_1, \lambda_2, \ldots, \lambda_r$ are small compared with the remainder, then values of $x_1, x_2, \ldots, x_k$ differing greatly from the correct solution but corresponding to points near the hyperplane h_r passing through the axes of $X_1, X_2, \ldots, X_r$, produce only small values of $\sum_{i=1}^{k} \delta_i^2$ and therefore only small values of $\delta_1, \delta_2, \ldots, \delta_k$ in (1). In this case, therefore, a wide variety of 'nearly correct' solutions of the equations $\mathbf{a}_0 + \mathbf{A}\mathbf{x} = 0$ exist leading to difficulties in the numerical solution, and the equations are said to be 'poorly conditioned'. An interesting example of such equations is given by Morris (1946). In the limit when $\lambda_1, \lambda_2, \ldots, \lambda_r$ are exactly zero, any point in the hyperplane h_r satisfies the equations, $\mathbf{A}$ being now of rank $k-r$, and this is the well-known case where r of the unknowns (that is, r co-ordinates, of a solution point) may be assigned arbitrarily.

Turing (1948) proposed as a measure of ill conditioning the quantity

$$c = k^{-1} N(\mathbf{A})\, N(\mathbf{A}^{-1}), \tag{12}$$

where $N(\mathbf{A}) = (\text{trace } \mathbf{A}'\mathbf{A})^{\frac{1}{2}} = (\sum_{ij} a_{ij}^2)^{\frac{1}{2}} = (\sum_{i} \lambda_i)^{\frac{1}{2}}$ is the norm of the matrix $\mathbf{A}$. The criterion is in line with the discussion above, for we see that

$$c = k^{-1}(\sum_i \lambda_1 \sum_i \lambda_i^{-1})^{\frac{1}{2}} \tag{13}$$

is a homogeneous function of degree zero in the λ's and is thus dependent only on their relative magnitudes. It takes the values of unity if all the λ's are equal (so that $\mathbf{A}$ is orthogonal), and is greater than unity otherwise. Its value will be large if any of the λ's are small compared with the others.

4. Examples of confidence regions

The confidence region for the solution of a set of linear equations would be expected to depend on (1) the magnitude of the errors in the coefficients and (2) the state of the conditioning of the equations. The separate contribution of these two influences can be seen particularly readily when the coefficients have equal variance and are uncorrelated, so that equation (7) may be used to obtain the confidence region.

For example, consider the pair of equations

$$3x_1 - 5x_2 - 1 = \delta_1, \tag{14}$$

$$3x_1 + 3x_2 - 9 = \delta_2, \tag{15}$$

when $\delta_1 = \delta_2 = 0$. These have the solution $x_1 = 2$, $x_2 = 1$ indicated by the cross in Fig. 1*a*. If the coefficients are subject to error, each with a known standard deviation equal to one, and they are uncorrelated, then the 95 % confidence region obtained by setting

$$v(a)\, s^2\, kF_\alpha(k, \phi) \quad \text{equal to} \quad 1 \times \chi^2_{0\cdot05}(2) = 5\cdot99$$

in equation (7) is shown by the bold line in Fig. 1*a*. We note from equation (7) that the boundary of the confidence region is the intersection of the surface

$$\delta_1^2 + \delta_2^2 = z \tag{16}$$

with the surface

$$(1 + x_1^2 + x_2^2)\, v(a)\, \sigma^2 \chi^2_\alpha = z. \tag{17}$$

Contours of these two surfaces for $z = 25$ and $z = 120$ are shown by the dotted lines in the figure and intersect on the confidence region. In separating equation (7) into the two parts we separate the two features which decide the confidence region. Equation (17) defines a saucer-like surface represented by the circular contours. The steepness of the surface depends on $v(a)\, \sigma^2$, the *variance of the coefficients*. Equation (16), on the other hand, defines the conditioning surface as an elliptical valley whose contours are concentric ellipses and whose attenuation is a measure of the *conditioning of the equations*. If $v(a)$ is small then the surface described by (17) will be shallow and will cut (16) to form a small region whose shape is very like that of the contours of the conditioning surface. If $v(a)$ is large, however, the surface (17) will rise more steeply and the confidence region will be larger and its shape will, to some extent, be distorted away from the shape of the contours of (16). In particular, the region will tend to extend farther from the solution point on the side remote from the origin.

Equations (14) and (15) are a well conditioned pair with $\lambda_1 = 16$, $\lambda_2 = 36$, $c = 1\cdot08$. A somewhat less conditioned pair is obtained if one-half of equation (14) is added to equation (15) to give a new equation to replace (14) so that we have

$$4\cdot5x_1 + 0\cdot5x_2 - 9\cdot5 = 0, \tag{18}$$

$$3\cdot0x_1 + 3\cdot0x_2 - 9\cdot0 = 0, \tag{19}$$

having the same solutions as before, but now $\lambda_1 = 4\cdot2$, $\lambda_2 = 34\cdot4$, $c = 1\cdot55$. Then if we again assume that the coefficients have standard deviations $\sigma = 1$ and are uncorrelated we obtain the open confidence interval shown by the bold lines in Fig. 1*b*.

The great difference in the shape of the confidence regions found with the two different pairs of equations is seen to be due to differences in conditioning as typified by the conditioning surfaces. Although the second set of equations is not seriously ill-conditioned, its conditioning surface is considerably more attenuated than that of the first set. Thus, when the surface corresponding to (17) is cut by that of (16) the open ended hyperbolic confidence region shown in Fig. 1*b* is obtained.

The effect of ill-conditioning is, in general, to cause the confidence region to spread out in the direction of the axes of the conditioning surface corresponding to the small latent roots of **A′A**.

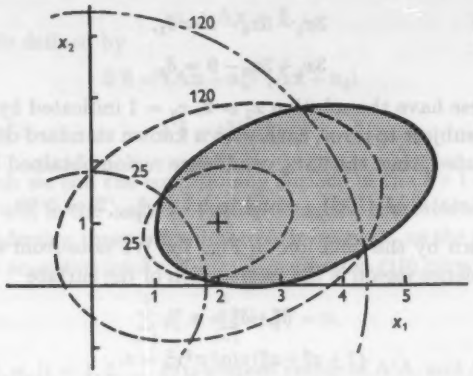


Fig. 1*a*. 95 % confidence region for the solution of the linear equations

$3x_1 - 5x_2 - 1 = 0$ and $3x_1 + 3x_2 - 9 = 0$.

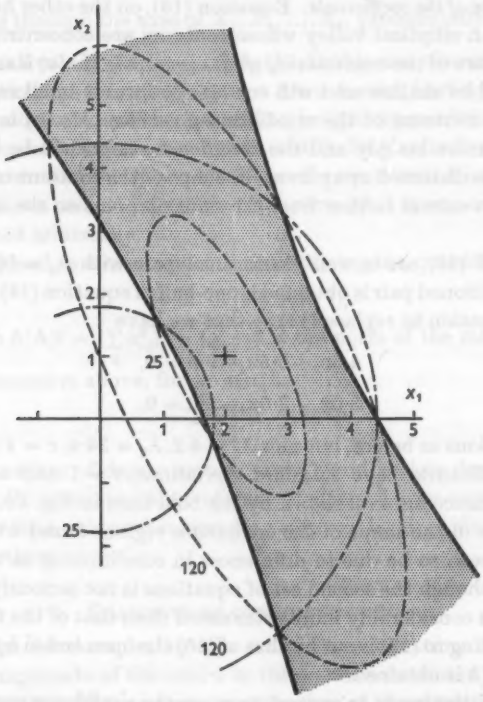


Fig. 1*b*. 95 % confidence region for the solution of the linear equations

$4{\cdot}5x_1 + 0{\cdot}5x_2 - 9{\cdot}5 = 0$ and $3{\cdot}0x_1 + 3{\cdot}0x_2 - 9{\cdot}0 = 0$.

5. The confidence region for the stationary point on a fitted second-degree surface

In a paper concerning experimental methods for attaining optimum conditions, Box & Wilson (1951) discussed the fitting by least squares of a polynomial equation relating an observed response to the levels of k independent variables, $x_1, x_2, \ldots, x_k$. Suggested experimental designs were discussed from the point of view of minimizing both the random errors in the estimates of the coefficients, and of the biases contributed by possible higher order terms which had been ignored in the response function.

Suppose that a polynomial equation of the second degree is fitted and for simplicity assume that there are only two independent variables x_1 and x_2. If the equation of the fitted response surface is

$$Y = a_{00} + a_{10}x_1 + a_{20}x_2 + \tfrac{1}{2}a_{11}x_1^2 + \tfrac{1}{2}a_{22}x_2^2 + a_{12}x_1x_2, \tag{20}$$

the centre of the fitted system is at a point corresponding to the solution of the equations

$$\frac{\partial Y}{\partial x_1} = a_{10} + a_{11}x_1 + a_{12}x_2 = \delta_1 = 0, \tag{21}$$

$$\frac{\partial Y}{\partial x_2} = a_{20} + a_{12}x_1 + a_{22}x_2 = \delta_2 = 0, \tag{22}$$

which are of the same form as equations (1) and (8) in the special case in which the matrix $\mathbf{A}$ is symmetric. Furthermore, the variance-covariance matrix of the least-squares estimates which corresponds to $\boldsymbol{\Omega}\sigma^2$ is known apart from σ^2 and is dependent only on the experimental design used, and an estimate s^2 of σ^2 is usually available either from the residual sum of squares (assuming an adequate model) or from some independent estimate obtained by replication. Thus to obtain a confidence region for the centre of the system we can immediately apply equation (5).

It should be remembered, of course, that in practice an important source of error not taken account of in the above confidence region arises due to the possible lack of fit of the second degree equation. Such lack of fit introduces errors not only directly in the sense that there is no second-degree equation which can adequately represent the surface, but also because the omission of higher order terms necessary to give an exact fit may cause the least-squares estimates of the second-degree equation to be biased. However, provided this limitation is borne in mind it is instructive to consider the size and type of confidence region that arises due to sampling errors in the coefficients alone.

We may first note the relationship that exists between the fitted response surface and the conditioning surface discussed in § 3. The equation of the fitted surface written in matrix notation is

$$Y - a_{00} = \mathbf{x}'\mathbf{a}_0 + \tfrac{1}{2}\mathbf{x}'\mathbf{A}\mathbf{x}, \tag{23}$$

and that of the conditioning surface is given by equation (9), where $\mathbf{A}$ is necessarily symmetric. On differentiation we see that if $\mathbf{A}$ is non-singular both systems have the same centre. Furthermore, the latent vectors of $\mathbf{A}'\mathbf{A} = \mathbf{A}^2$ are the same as those of $\mathbf{A}$ and the latent roots are the squares of those of $\mathbf{A}$.

Thus the centre and the principal axes of the conditioning surface are the same as those of the response surface, but attenuation of the response surface is accompanied by even greater attenuation of the conditioning surface and the contours of the conditioning surface are always ellipses. We see, therefore, that to the extent to which the confidence region reflects the influence of the conditioning surface, it will (to some lesser extent) reflect the characteristics of the response surface itself.

It should be remembered that the method described provides a confidence region for a stationary point, which has as its co-ordinates the solution to the equations (21) and (22). If the coefficients in (21) and (22) are such as to indicate that the fitted surface possesses a maximum, *and* the confidence region is closed, we may assert that this is a confidence region for a true maximum (that is to say, not merely a confidence region for a stationary point). This is so because, assuming that the surface can be represented by an equation of second degree, if the plausible range of variation in the a's includes surfaces in which one or more of the λ's could change in sign, then points lying along the axes corresponding to these roots would be acceptable as solutions of equations (21) and (22), and open contours would necessarily result.

6. Example of a confidence region obtained using an experimental design

To give some appreciation of the type of confidence regions for an experimentally determined stationary point met in practice, a statistical experiment was carried out as follows. It was assumed that the true equation of a surface was

$$\eta = 78{\cdot}373 + 4{\cdot}533x_1 - 1{\cdot}867x_2 - 3{\cdot}333x_1^2 - 3{\cdot}333x_2^2 + 4{\cdot}000x_1x_2, \tag{24}$$

and that a 3^2 factorial experiment was conducted at the levels $-1, 0, 1$ for the independent variables x_1 and x_2. Random normal deviates with $\sigma = 1$ were added to the values of η at the nine points in the design. The resulting values are recorded in Fig. 2. The second-degree equation fitted by least squares to these points was

$$Y = 77{\cdot}82 + 3{\cdot}70x_1 - 1{\cdot}62x_2 - 2{\cdot}83x_1^2 - 2{\cdot}58x_2^2 + 4{\cdot}02x_1x_2, \tag{25}$$

whose contours are indicated by the dotted lines in Fig. 2 and whose centre, $x_1^0 = 0{\cdot}96$, $x_2^0 = 0{\cdot}44$, is indicated by a cross. The residual sum of squares, based on 3 degrees of freedom was 7·20, giving an estimate of $s^2 = 2{\cdot}40$, assuming the model is adequate. The variances and covariances for the coefficients of (25) estimated from a 3^2 factorial are

$$V(a_{00}) = \tfrac{5}{9}\sigma^2, \quad V(a_{10}) = V(a_{20}) = \tfrac{1}{6}\sigma^2, \quad V(a_{11}) = V(a_{22}) = 2\sigma^2, \quad V(a_{12}) = \tfrac{1}{4}\sigma^2,$$
$$\operatorname{cov}(a_{00}, a_{11}) = \operatorname{cov}(a_{00}, a_{22}) = -\tfrac{2}{3}\sigma^2.$$

Substituting the values of $s^2 = 2{\cdot}40$, and the value of $F_{0{\cdot}05}(2, 3) = 9{\cdot}55$ in equation (5), we obtain for the 95 % confidence region the shaded portion indicated in Fig. 2.

In practice, having fitted the second-degree surface the experimenter would perform confirmatory experiments. A natural location for additional experiments would be along the principal axes of the fitted conic. We have assumed that six further points were added as shown in Fig. 3. The values which might have been obtained at these points, indicated in Fig. 3, were secured by calculating the values at the points using equation (24) and adding random normal deviates as before. Using all 15 points the equation was refitted giving

$$Y = 77{\cdot}95 + 3{\cdot}76x_1 - 1{\cdot}57x_2 - 2{\cdot}87x_1^2 - 2{\cdot}64x_2^2 + 3{\cdot}84x_1x_2. \tag{26}$$

The symmetric variance-covariance matrix for the new estimates a_{ij} is indicated below:

$$\begin{array}{c} \begin{matrix} 00 & 10 & 20 & 11 & 22 & 12 \end{matrix} \\ \begin{bmatrix} 0{\cdot}18684 & -0{\cdot}00296 & 0{\cdot}00921 & -0{\cdot}10282 & -0{\cdot}23162 & 0{\cdot}08576 \\ & 0{\cdot}13091 & -0{\cdot}01361 & -0{\cdot}12340 & 0{\cdot}05762 & -0{\cdot}01263 \\ & & 0{\cdot}13574 & 0{\cdot}01184 & 0{\cdot}08260 & -0{\cdot}02844 \\ & & & 0{\cdot}43848 & -0{\cdot}15168 & -0{\cdot}09496 \\ & & & & 0{\cdot}95532 & -0{\cdot}18868 \\ & & & & & 0{\cdot}14688 \end{bmatrix} \sigma^2. \end{array} \tag{27}$$

The residual sum of squares was 9·60, and hence the residual error variance s^2 was 1·07, now based on 9 degrees of freedom. The centre of the newly fitted equation is at the point $x_1 = 0{\cdot}89$, $x_2^0 = 0{\cdot}37$ indicated by the cross in Fig. 3.

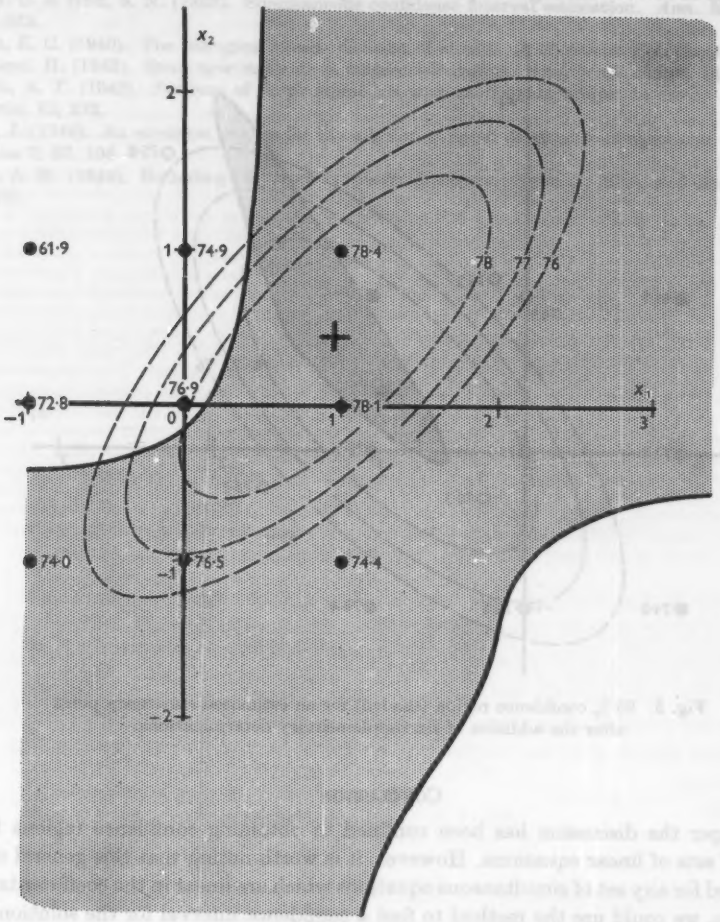


Fig. 2. 95 % confidence region (shaded) for an estimated stationary point based on nine determinations from a 3^2 factorial design.

The confidence region would now be expected to be considerably smaller

(i) because of the influence of the extra points,

(ii) because of the larger number of degrees of freedom upon which F is based (the critical value of F is reduced from 9·55 to 4·26),

(iii) the first estimate of σ^2 ($s^2 = 2{\cdot}4$) happens to be considerably greater than the second estimate ($s^2 = 1{\cdot}07$).

The new confidence region, shown as the shaded area in Fig. 3, is closed but attenuated.

The size and attenuation of the regions found emphasize the considerable uncertainty that would sometimes exist in the location of a maximum, even when the profound effect of lack of fit of the second-degree equation was ignored.

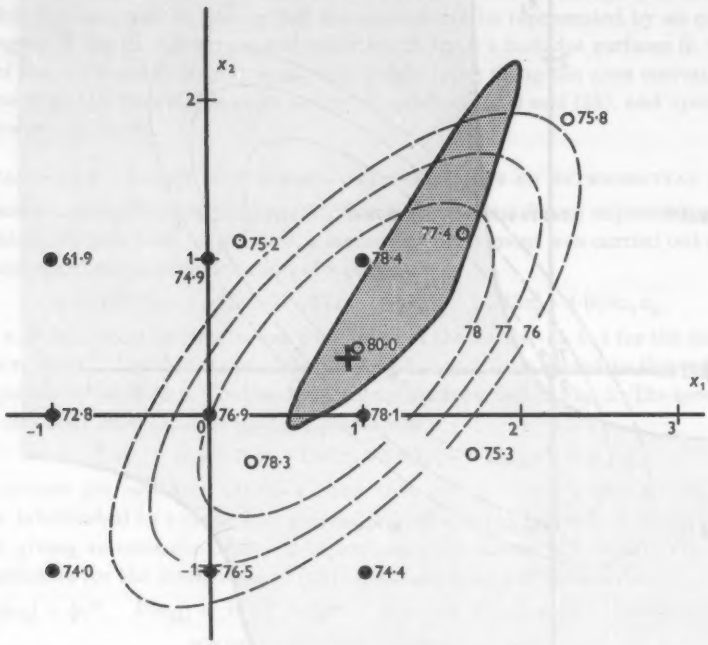


Fig. 3. 95 % confidence region (shaded) for an estimated stationary point after the addition of six supplementary determinations.

Conclusion

In this paper the discussion has been confined to obtaining confidence regions for the solution of sets of linear equations. However, it is worth noting that this general method may be used for any set of simultaneous equations which are linear in the coefficients. Thus, in principle, we could use the method to find a confidence interval for the solution of the k equations

$$a_{i0} + \sum_{j=1}^{k} a_{ij} f_j(x_1, x_2, \ldots, x_k) = 0 \quad (i = 1, 2, \ldots, k). \tag{28}$$

Consequently, confidence intervals could be obtained for a stationary point on a surface represented by any equation linear in the coefficients, and not only for the quadric surface we have discussed.

REFERENCES

Box, G. E. P. & Wilson, K. B. (1951). On the experimental attainment of optimum conditions. *J. R. Statist. Soc.* B, **13**, 1.

Bose, R. C. & Roy, S. N. (1953). Simultaneous confidence interval estimation. *Ann. Math. Statist.* **24**, 513.

Fieller, E. C. (1940). The biological standardization of insulin. *J. R. Statist. Soc. Suppl.* **7**, 1.

Hotelling, H. (1943). Some new methods in matrix calculation. *Ann. Math. Statist.* **14**, 1.

Lonseth, A. T. (1942). Systems of linear equations with coefficients subject to error. *Ann. Math. Statist.* **13**, 332.

Morris, J. (1946). An escalator process for the solution of linear simultaneous equations. *Phil. Mag. Series* **7**, **37**, 106.

Turing, A. M. (1948). Rounding off errors in matrix processes. *Quart. J. Mech. and Applied Math.* **1**, 287.

ON THE MOMENTS OF ORDER STATISTICS IN SAMPLES FROM NORMAL POPULATIONS

By H. RUBEN

Agricultural Research Council Fellow

The purpose of this paper is to show the geometrical significance of the moments of order statistics derived from normal populations. It appears that these, as well as the moment-generating function of the square of any order statistic, are intimately related to the contents of the members of a class of hyperspherical simplices. Further geometrical interpretations, together with the extension of the present results to bivariate moments and other properties, will be provided in subsequent publications.

The problem of order statistics in normal populations has been extensively considered in the literature. For example,† Tippett (1925) gives the second, third and fourth moments of the extreme order statistics for a few sample sizes. Hojo (1931) examines the sampling variation of the median, quartiles and interquartile distance in samples from normal populations and computes a large number of integrals for this purpose. Cole (1951) produces a very simple recurrence relationship between the 'normalized' moments which enables the normalized moments for all samples of size no greater than n to be obtained, by successive differencing, from the 'normalized' moments of the extreme order statistics in samples of size $m \leqslant n$.‡ Hastings, Mosteller, Tukey & Winsor (1947) give, among other results, the means, variances, covariances and correlations of order statistics in samples of ten or less from a normal population, some of the results being given to only two decimal places because of the extreme labour of the computation. Jones (1948) and Godwin (1949*a*, *b*) have obtained exact values for some of the lower moments. In particular, Godwin (1949*a*, p. 283) proceeds in a more systematic manner, and approaches most closely the essential idea upon which this paper is based.

In general, it may be said that many statisticians have attacked the problem in a rather disjointed and fragmentary manner, usually from the small sample end, but have failed to develop a systematic attack which shall at the same time throw light on the interconnexion between the moments and enable the computation of the moments to become an economic proposition. It is believed that these conditions are met by this paper.

1. The reduction of a certain generalized class of integrals

We shall here consider the properties of a class of integrals as defined in equation (1). This equation gives the general form of integral§ needed in the derivation of the series (see equations (65) and (66)) for the moments of the order statistics. In this section we shall show how any member of the class of integrals can be expressed as the series (23) depending on functions $\phi(0; \alpha; \beta, \gamma)$ and containing coefficients whose values are listed at the end of the section.

Let then
$$\phi(s; \alpha; \beta, \gamma) = \int_{-\infty}^{\infty} x^s z^\alpha p^\beta q^\gamma \, dx, \tag{1}$$

where
$$z = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}, \tag{2}$$

$$p = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{x} e^{-\frac{1}{2}\xi^2} \, d\xi, \tag{3}$$

$$q = \frac{1}{\sqrt{(2\pi)}} \int_{x}^{\infty} e^{-\frac{1}{2}\xi^2} \, d\xi, \tag{4}$$

and s, β and γ are non-negative integers, while α is real and positive.

† The references given are not intended to provide an exhaustive bibliography.

‡ The author has found that the cumulative error induced by the differencing increases too rapidly for the moments of the order statistics, except for those which are only a few places removed from the extreme, to be computed with a reasonable degree of accuracy.

§ Such an integral, with $s = 0$ and $\alpha \neq 1$, arises also in connexion with the moment generating function of the square of the general order statistic in samples from normal populations (see equation (51)).

On integration by parts,

$$\phi(s;\alpha;\beta,\gamma) = -\int_{-\infty}^{\infty} x^{s-1}z^{\alpha-1}p^{\beta}q^{\gamma}\frac{dz}{dx}dx$$

$$= \int_{-\infty}^{\infty} z\frac{d}{dx}(x^{s-1}z^{\alpha-1}p^{\beta}q^{\gamma})\,dx$$

$$= (s-1)\int_{-\infty}^{\infty} x^{s-2}z^{\alpha}p^{\beta}q^{\gamma}\,dx-(\alpha-1)\int_{-\infty}^{\infty} x^{s}z^{\alpha}p^{\beta}q^{\gamma}\,dx$$

$$+\beta\int_{-\infty}^{\infty} x^{s-1}z^{\alpha+1}p^{\beta-1}q^{\gamma}\,dx-\gamma\int_{-\infty}^{\infty} x^{s-1}z^{\alpha+1}p^{\beta}q^{\gamma-1}dx \quad (s=1,2,\ldots).$$

Hence
$$\phi(s;\alpha;\beta,\gamma) = \frac{s-1}{\alpha}\phi(s-2;\alpha;\beta,\gamma)+\frac{\beta}{\alpha}\phi(s-1;\alpha+1;\beta-1,\gamma)$$
$$-\frac{\gamma}{\alpha}\phi(s-1;\alpha+1;\beta,\gamma-1) \quad (s=1,2,\ldots). \tag{5}$$

Let **A** denote the operator which *increases* α by 1 and **B** and $\boldsymbol{\Gamma}$ the operators which *decrease* β and γ, respectively, by 1. It is useful also to introduce the associated operators **P**, **H**, **K** and **L** defined by

$$\mathbf{P} \equiv \frac{1}{\alpha}\mathbf{A}, \tag{6}$$

$$\mathbf{H} \equiv \beta\mathbf{B}, \tag{7}$$

$$\mathbf{K} \equiv \gamma\boldsymbol{\Gamma}, \tag{8}$$

$$\mathbf{L} \equiv \mathbf{H}-\mathbf{K}. \tag{9}$$

Certain rather obvious properties of these operators, as well as of functions of the operators, which will be used in the sequel should be noted in passing. Any two operators which affect, and therefore bind, different and disjoint sets of arguments commute with each other. Thus **B** commutes with $\boldsymbol{\Gamma}$, $\mathbf{H}^m$ with $\mathbf{K}^n$ $(m,n=0,1,2,\ldots)$, **P** with **L**, $\mathbf{P}^m$ with $\mathbf{L}^n$, etc. We state some additional properties of the operators. First,

$$\mathbf{P}^m \equiv \frac{1}{\alpha^{-(m)}}\mathbf{A}^m \quad (m=0,1,2,\ldots), \tag{10}$$

$$\left.\begin{aligned}\mathbf{H}^m &\equiv \beta^{(m)}\mathbf{B}^m\\ \mathbf{K}^m &\equiv \gamma^{(m)}\boldsymbol{\Gamma}^m\end{aligned}\right\} \quad (m=0,1,2,\ldots), \tag{11}$$

where $u^{-(m)}$ and $u^{(m)}$ denote the ascending and descending factorials, respectively, of degree m in u, i.e.

$$\left.\begin{aligned}u^{-(0)} &= 1,\\ u^{-(m)} &= u(u+1)\ldots(u+m-1) \quad (m=1,2,\ldots),\end{aligned}\right\} \tag{12}$$

$$\left.\begin{aligned}u^{(0)} &= 1,\\ u^{(m)} &= u(u-1)\ldots(u-m+1) \quad (m=1,2,\ldots),\end{aligned}\right\} \tag{13}$$

and $\mathbf{A}^0$, $\mathbf{B}^0$, $\boldsymbol{\Gamma}^0$, etc., are defined as identity operators. The validity of the identities in (10) and (11) follows by induction. For

$$\mathbf{P}\frac{1}{\alpha^{-(m)}}\mathbf{A}^m = \frac{1}{\alpha}\frac{1}{(\alpha+1)^{-(m)}}\mathbf{A}^{m+1} = \frac{1}{\alpha^{-(m+1)}}\mathbf{A}^{m+1},$$

and since (10) is true for $m = 0$, it is identically true for all non-negative integral m. Likewise (11) is valid for all such values of m, since

$$\mathbf{H}\beta^{(m)}\mathbf{B}^m = \beta(\beta-1)^{(m)}\,\mathbf{B}^{m+1} = \beta^{(m+1)}\mathbf{B}^{m+1}.$$

From (10) and (11) it follows directly that **P**, **H** and **K** are governed by the index law:

$$\mathbf{P}^m\mathbf{P}^n \equiv \mathbf{P}^{m+n} \equiv \mathbf{P}^n\mathbf{P}^m \qquad (m,n = 0,1,2,\ldots), \tag{14}$$

$$\left.\begin{aligned}\mathbf{H}^m\mathbf{H}^n &\equiv \mathbf{H}^{m+n} \equiv \mathbf{H}^n\mathbf{H}^m\\ \mathbf{K}^m\mathbf{K}^n &\equiv \mathbf{K}^{m+n} \equiv \mathbf{K}^n\mathbf{K}^m\end{aligned}\right\} \quad (m,n = 0,1,2,\ldots). \tag{15}$$

Secondly,

$$\mathbf{L}^m \equiv \sum_{j=0}^{m}(-)^j\binom{m}{j}\mathbf{K}^j\mathbf{H}^{m-j} \equiv \sum_{j=0}^{m}(-)^j\binom{m}{j}\gamma^{(j)}\beta^{(m-j)}\mathbf{L}^j\mathbf{\Gamma}^{m-j} \quad (m = 0,1,2,\ldots), \tag{16}$$

i.e. $\mathbf{L}^m$ may be obtained by the formal bionomial expansion of $(\mathbf{H}-\mathbf{K})^m$. The relationship (16) is again proved by induction. The proof will not be given here, being identical in form with the text-book proof of the binomial expansion for positive integral indices, when the elements are constants and not operators. Again

$$\mathbf{L}^m\mathbf{L}^n \equiv \mathbf{L}^{m+n} \equiv \mathbf{L}^n\mathbf{L}^m \quad (m,n = 0,1,2,\ldots). \tag{17}$$

For
$$\begin{aligned}\mathbf{L}^m\mathbf{L}^n &= \sum_{i=0}^{m}(-)^i\binom{m}{i}\mathbf{K}^i\mathbf{H}^{m-i}\sum_{j=0}^{n}(-)^j\binom{n}{j}\mathbf{K}^j\mathbf{H}^{n-j}\\ &= \sum_{i=0}^{m}\sum_{j=0}^{n}(-)^{i+j}\binom{m}{i}\binom{n}{j}\mathbf{K}^{i+j}\mathbf{H}^{m+n-i-j}\\ &= \sum_{r=0}^{m+n}\sum_{i+j=r}(-)^r\binom{m}{i}\binom{n}{j}\mathbf{K}^r\mathbf{H}^{m+n-r}\\ &= \sum_{r=0}^{m+n}(-)^r\binom{m+n}{r}\mathbf{K}^r\mathbf{H}^{m+n-r}.\end{aligned}$$

Similarly,
$$\mathbf{L}^n\mathbf{L}^m = \sum_{r=0}^{m+n}(-)^r\binom{m+n}{r}\mathbf{K}^r\mathbf{H}^{m+n-r},$$

and hence (17) is proved.

Finally, **L** and the operation of integration commute. Specifically,

$$\mathbf{L}\int_{\theta_0}^{\theta} f(\theta';\beta,\gamma)\,d\theta' \equiv \int_{\theta_0}^{\theta}[\mathbf{L}f(\theta';\beta,\gamma)]\,d\theta'. \tag{18}$$

For, let
$$F(\theta;\beta,\gamma) = \int_{\theta_0}^{\theta} f(\theta';\beta,\gamma)\,d\theta'.$$

Then
$$\begin{aligned}\mathbf{L}\int_{\theta_0}^{\theta} f(\theta';\beta,\gamma)\,d\theta' &= (\beta\mathbf{B}-\gamma\mathbf{\Gamma})\,F(\theta;\beta,\gamma)\\ &= \beta F(\theta;\beta-1,\gamma)-\gamma F(\theta;\beta,\gamma-1)\\ &= \int_{\theta_0}^{\theta}[\mathbf{L}f(\theta';\beta,\gamma)]\,d\theta'.\end{aligned}$$

We note from (16) that if ω is independent of β and γ, then

$$\mathbf{L}^m\omega = \omega \sum_{j=0}^{m} (-)^j \binom{m}{j} \gamma^{(j)}\beta^{(m-j)}, \tag{19}$$

and that if β and γ are non-negative integers with $m > \beta+\gamma$, m being integral, then $\mathbf{L}^m \equiv 0$:

$$\mathbf{L}^m \equiv 0 \quad (m = \beta+\gamma+1, \beta+\gamma+2, \ldots; \beta, \gamma = 0, 1, 2, \ldots). \tag{20}$$

Also, from (18),

$$\mathbf{L}^m \int_{\theta_0}^{\theta} f(\theta'; \beta, \gamma)\, d\theta' \equiv \int_{\theta_0}^{\theta} [\mathbf{L}^m f(\theta'; \beta, \gamma)]\, d\theta' \quad (m = 0, 1, 2, \ldots). \tag{21}$$

Reverting to equation (5), this may be written symbolically in the form

$$\phi(s; \alpha; \beta, \gamma) = \frac{s-1}{\alpha}\phi(s-2; \alpha; \beta, \gamma) + \mathbf{PL}\phi(s-1; \alpha; \beta, \gamma) \quad (s = 1, 2, \ldots). \tag{22}$$

The fundamental reduction formula (22) will be used to reduce $\phi(s; \alpha; \beta, \gamma)$, when s is a positive integer, to a linear combination of ϕ-functions with zero as the first argument. In fact we shall show by induction that the solution of the recurrence equation (22) is expressed by

$$\left.\begin{aligned} \phi(2k+1; \alpha; \beta, \gamma) &= \sum_{i=0}^{k} a_{2k+1,\,2i+1}(\alpha)(\mathbf{PL})^{2i+1}\phi(0; \alpha; \beta, \gamma), \\ \phi(2k; \alpha; \beta, \gamma) &= \sum_{i=0}^{k} a_{2k,\,2i}(\alpha)(\mathbf{PL})^{2i}\phi(0; \alpha; \beta, \gamma). \end{aligned}\right\} \quad (k = 0, 1, 2, \ldots), \tag{23}$$

provided the a's are appropriately chosen. (The value $k = 0$ is included for the sake of completeness.)

Let it be assumed that (23) is true for some non-negative integer k. On using this assumption together with (22),

$$\begin{aligned} \phi(2k+2; \alpha; \beta, \gamma) &= \frac{2k+1}{\alpha}\phi(2k; \alpha; \beta, \gamma) + \mathbf{PL}\phi(2k+1; \alpha; \beta, \gamma) \\ &= \frac{2k+1}{\alpha}\sum_{i=0}^{k} a_{2k,\,2i}(\alpha)\,(\mathbf{PL})^{2i}\phi(0; \alpha; \beta, \gamma) \\ &\qquad + \mathbf{PL}\sum_{i=0}^{k} a_{2k+1,\,2i+1}(\alpha)\,(\mathbf{PL})^{2i+1}\phi(0; \alpha; \beta, \gamma) \\ &= \frac{2k+1}{\alpha}\sum_{i=0}^{k} a_{2k,\,2i}(\alpha)\,(\mathbf{PL})^{2i}\phi(0; \alpha; \beta, \gamma) \\ &\qquad + \sum_{i=0}^{k} a_{2k+1,\,2i+1}(\alpha+1)\,(\mathbf{PL})^{2i+2}\phi(0; \alpha; \beta, \gamma) \\ &= \sum_{i=0}^{k+1} a_{2k+2,\,2i}(\alpha)\,(\mathbf{PL})^{2i}\phi(0; \alpha; \beta, \gamma), \end{aligned}$$

provided

$$a_{2k+2,\,2i}(\alpha) = \frac{2k+1}{\alpha}a_{2k,\,2i}(\alpha) + a_{2k+1,\,2i+1}(\alpha+1) \quad (i = 1, 2, \ldots, k), \tag{24}$$

$$a_{2k+2,\,0}(\alpha) = \frac{2k+1}{\alpha}a_{2k,\,0}(\alpha), \tag{25}$$

$$a_{2k+2,\,2k+2}(\alpha) = a_{2k+1,\,2k+1}(\alpha+1). \tag{26}$$

Similarly, and using the result just established,

$$\begin{aligned}\phi(2k+3;\alpha;\beta,\gamma) &= \frac{2k+2}{\alpha}\phi(2k+1;\alpha;\beta,\gamma)+\mathbf{PL}\phi(2k+2;\alpha;\beta,\gamma)\\ &= \frac{2k+2}{\alpha}\sum_{i=0}^{k} a_{2k+1,2i+1}(\alpha)\,(\mathbf{PL})^{2i+1}\phi(0;\alpha;\beta,\gamma)\\ &\qquad +\mathbf{PL}\sum_{i=0}^{k+1} a_{2k+2,2i}(\alpha)\,(\mathbf{PL})^{2i}\phi(0;\alpha;\beta,\gamma)\\ &= \frac{2k+2}{\alpha}\sum_{i=0}^{k} a_{2k+1,2i+1}(\alpha)\,(\mathbf{PL})^{2i+1}\phi(0;\alpha;\beta,\gamma)\\ &\qquad +\sum_{i=0}^{k+1} a_{2k+2,2i}(\alpha+1)\,(\mathbf{PL})^{2i+1}\phi(0;\alpha;\beta,\gamma)\\ &= \sum_{i=0}^{k+1} a_{2k+3,2i+1}(\alpha)\,(\mathbf{PL})^{2i+1}\phi(0;\alpha;\beta,\gamma),\end{aligned}$$

provided

$$a_{2k+3,2i+1}(\alpha) = \frac{2k+2}{\alpha}a_{2k+1,2i+1}(\alpha)+a_{2k+2,2i}(\alpha+1)\quad(i=0,1,2,\ldots,k), \tag{27}$$

$$a_{2k+3,2k+3}(\alpha) = a_{2k+2,2k+2}(\alpha+1). \tag{28}$$

It follows that if (23) is valid for some non-negative integer k, it is valid also for $k+1$, provided the a's satisfy equations (24)–(28). Consequently (23) is valid for all non-negative integral k, being obviously true for $k=0$ with $a_{0,0}(\alpha)=1$ and $a_{1,1}(\alpha)=1$.

The equations (24)–(28) may be solved for the a's quite systematically. In particular, by successive application of (25),

$$a_{2k+2,0}(\alpha) = 1.3.5.\ldots.(2k+1)/\alpha^{k+1}. \tag{29}$$

Again, by successive application of (26) and (28),

$$a_{2k+3,2k+3}(\alpha) = a_{2k+2,2k+2}(\alpha+1) = \ldots = a_{0,0}(\alpha+2k+3) = 1. \tag{30}$$

Similarly,

$$a_{2k+2,2k+2}(\alpha) = 1. \tag{30'}$$

Setting $i=k$ in (24) and using (27), (30) and (30'),

$$\begin{aligned}a_{2k+2,2k}(\alpha) &= \frac{2k+1}{\alpha}a_{2k,2k}(\alpha)+a_{2k+1,2k-1}(\alpha+1)\\ &= \frac{2k+1}{\alpha}+a_{2k+1,2k-1}(\alpha+1)\\ &= \frac{2k+1}{\alpha}+\frac{2k}{\alpha+1}+a_{2k,2k-2}(\alpha+2) = \ldots\\ &= \frac{2k+1}{\alpha}+\frac{2k}{\alpha+1}+\ldots+\frac{2}{\alpha+2k-1}+\frac{1}{\alpha+2k}.\end{aligned} \tag{31}$$

Similarly, setting $i=k$ in (27) and using (30) and (31),

$$a_{2k+3,2k+1}(\alpha) = \frac{2k+2}{\alpha}+\frac{2k+1}{\alpha+1}+\ldots+\frac{2}{\alpha+2k}+\frac{1}{\alpha+2k+1}. \tag{31'}$$

The equations (29), (30), (30'), (31) and (31') give the forms which the a's take in particular cases. In general, however, the structure of the coefficients is more complicated, but they

can be determined systematically by successive application of (24) and (27). The values of the coefficients relating to the various values of s up to and including $s = 10$ are given below:

$s = 0$:

$$a_{0,0}(\alpha) = 1$$

$s = 1$:

$$a_{1,1}(\alpha) = 1$$

$s = 2$:

$$a_{2,0}(\alpha) = \frac{1}{\alpha}$$

$$a_{2,2}(\alpha) = 1$$

$s = 3$:

$$a_{3,1}(\alpha) = \frac{2}{\alpha} + \frac{1}{\alpha+1}$$

$$a_{3,3}(\alpha) = 1$$

$s = 4$:

$$a_{4,0}(\alpha) = 1$$

$$a_{4,2}(\alpha) = \frac{3}{\alpha} + \frac{2}{\alpha+1} + \frac{1}{\alpha+2}$$

$$a_{4,4}(\alpha) = 1$$

$s = 5$:

$$a_{5,1}(\alpha) = \frac{4}{\alpha}\left(\frac{2}{\alpha} + \frac{1}{\alpha+1}\right) + \frac{3}{(\alpha+1)^2}$$

$$a_{5,3}(\alpha) = \frac{4}{\alpha} + \frac{3}{\alpha+1} + \frac{2}{\alpha+2} + \frac{1}{\alpha+3}$$

$$a_{5,5}(\alpha) = 1$$

$s = 6$:

$$a_{6,0}(\alpha) = \frac{15}{\alpha^3}$$

$$a_{6,2}(\alpha) = \frac{5}{\alpha}\left(\frac{3}{\alpha} + \frac{2}{\alpha+1} + \frac{1}{\alpha+2}\right) + \frac{4}{\alpha+1}\left(\frac{2}{\alpha+1} + \frac{1}{\alpha+2}\right) + \frac{3}{(\alpha+2)^2}$$

$$a_{6,4}(\alpha) = \frac{5}{\alpha} + \frac{4}{\alpha+1} + \frac{3}{\alpha+2} + \frac{2}{\alpha+3} + \frac{1}{\alpha+4}$$

$$a_{6,6}(\alpha) = 1$$

$s = 7$:

$$a_{7,1}(\alpha) = \frac{6}{\alpha}\left(\frac{4}{\alpha}\left(\frac{2}{\alpha}+\frac{1}{\alpha+1}\right)+\frac{3}{(\alpha+1)^2}\right)+\frac{15}{(\alpha+1)^3}$$

$$a_{7,3}(\alpha) = \frac{6}{\alpha}\left(\frac{4}{\alpha}+\frac{3}{\alpha+1}+\frac{2}{\alpha+2}+\frac{1}{\alpha+3}\right)+\frac{5}{\alpha+1}\left(\frac{3}{\alpha+1}+\frac{2}{\alpha+2}+\frac{1}{\alpha+3}\right)$$
$$+\frac{4}{\alpha+2}\left(\frac{2}{\alpha+2}+\frac{1}{\alpha+3}\right)+\frac{3}{(\alpha+3)^2}$$

$$a_{7,5}(\alpha) = \frac{6}{\alpha}+\frac{5}{\alpha+1}+\frac{4}{\alpha+2}+\frac{3}{\alpha+3}+\frac{2}{\alpha+4}+\frac{1}{\alpha+5}$$

$$a_{7,7}(\alpha) = 1$$

$s = 8$:

$$a_{8,0}(\alpha) = \frac{105}{\alpha^4}$$

$$a_{8,2}(\alpha) = \frac{7}{\alpha}\left(\frac{5}{\alpha}\left(\frac{3}{\alpha}+\frac{2}{\alpha+1}+\frac{1}{\alpha+2}\right)+\frac{4}{\alpha+1}\left(\frac{2}{\alpha+1}+\frac{1}{\alpha+2}\right)+\frac{3}{(\alpha+2)^2}\right)$$
$$+\frac{6}{\alpha+1}\left(\frac{4}{\alpha+1}\left(\frac{2}{\alpha+1}+\frac{1}{\alpha+2}\right)+\frac{3}{(\alpha+2)^2}\right)+\frac{15}{(\alpha+2)^3}$$

$$a_{8,4}(\alpha) = \frac{7}{\alpha}\left(\frac{5}{\alpha}+\frac{4}{\alpha+1}+\frac{3}{\alpha+2}+\frac{2}{\alpha+3}+\frac{1}{\alpha+4}\right)+\frac{6}{\alpha+1}\left(\frac{4}{\alpha+1}+\frac{3}{\alpha+2}+\frac{2}{\alpha+3}+\frac{1}{\alpha+4}\right)$$
$$+\frac{5}{\alpha+2}\left(\frac{3}{\alpha+2}+\frac{2}{\alpha+3}+\frac{1}{\alpha+4}\right)+\frac{4}{\alpha+3}\left(\frac{2}{\alpha+3}+\frac{1}{\alpha+4}\right)+\frac{3}{(\alpha+4)^2}$$

$$a_{8,6}(\alpha) = \frac{7}{\alpha}+\frac{6}{\alpha+1}+\frac{5}{\alpha+2}+\frac{4}{\alpha+3}+\frac{3}{\alpha+4}+\frac{2}{\alpha+5}+\frac{1}{\alpha+6}$$

$$a_{8,8}(\alpha) = 1$$

$s = 9$:

$$a_{9,1}(\alpha) = \frac{8}{\alpha}\left(\frac{6}{\alpha}\left(\frac{4}{\alpha}\left(\frac{2}{\alpha}+\frac{1}{\alpha+1}\right)+\frac{3}{(\alpha+1)^2}\right)+\frac{15}{(\alpha+1)^3}\right)+\frac{105}{(\alpha+1)^4}$$

$$a_{9,3}(\alpha) = \frac{8}{\alpha}\left(\frac{6}{\alpha}\left(\frac{4}{\alpha}+\frac{3}{\alpha+1}+\frac{2}{\alpha+2}+\frac{1}{\alpha+3}\right)+\frac{5}{\alpha+1}\left(\frac{3}{\alpha+1}+\frac{2}{\alpha+2}+\frac{1}{\alpha+3}\right)\right.$$
$$\left.+\frac{4}{\alpha+2}\left(\frac{2}{\alpha+2}+\frac{1}{\alpha+3}\right)+\frac{3}{(\alpha+3)^2}\right)$$
$$+\frac{7}{\alpha+1}\left(\frac{5}{\alpha+1}\left(\frac{3}{\alpha+1}+\frac{2}{\alpha+2}+\frac{1}{\alpha+3}\right)+\frac{4}{\alpha+2}\left(\frac{2}{\alpha+2}+\frac{1}{\alpha+3}\right)+\frac{3}{(\alpha+3)^2}\right)$$
$$+\frac{6}{\alpha+2}\left(\frac{4}{\alpha+2}\left(\frac{2}{\alpha+2}+\frac{1}{\alpha+3}\right)+\frac{3}{(\alpha+3)^2}\right)+\frac{15}{(\alpha+3)^3}$$

$$a_{9,5}(\alpha) = \frac{8}{\alpha}\left(\frac{6}{\alpha}+\frac{5}{\alpha+1}+\frac{4}{\alpha+2}+\frac{3}{\alpha+3}+\frac{2}{\alpha+4}+\frac{1}{\alpha+5}\right)$$
$$+\frac{7}{\alpha+1}\left(\frac{5}{\alpha+1}+\frac{4}{\alpha+2}+\frac{3}{\alpha+3}+\frac{2}{\alpha+4}+\frac{1}{\alpha+5}\right)+\frac{6}{\alpha+2}\left(\frac{4}{\alpha+2}+\frac{3}{\alpha+3}+\frac{2}{\alpha+4}+\frac{1}{\alpha+5}\right)$$
$$+\frac{5}{\alpha+3}\left(\frac{3}{\alpha+3}+\frac{2}{\alpha+4}+\frac{1}{\alpha+5}\right)+\frac{4}{\alpha+4}\left(\frac{2}{\alpha+4}+\frac{1}{\alpha+5}\right)+\frac{3}{(\alpha+5)^2}$$

$s = 9$:

$$a_{9,7}(\alpha) = \frac{8}{\alpha} + \frac{7}{\alpha+1} + \frac{6}{\alpha+2} + \frac{5}{\alpha+3} + \frac{4}{\alpha+4} + \frac{3}{\alpha+5} + \frac{2}{\alpha+6} + \frac{1}{\alpha+7}$$

$$a_{9,9}(\alpha) = 1$$

$s = 10$:

$$a_{10,0}(\alpha) = \frac{945}{\alpha^5}$$

$$a_{10,2}(\alpha) = \frac{9}{\alpha}\left(\frac{7}{\alpha}\left(\frac{5}{\alpha}\left(\frac{3}{\alpha} + \frac{2}{\alpha+1} + \frac{1}{\alpha+2}\right) + \frac{4}{\alpha+1}\left(\frac{2}{\alpha+1} + \frac{1}{\alpha+2}\right) + \frac{3}{(\alpha+2)^2}\right)\right.$$
$$+ \frac{6}{\alpha+1}\left(\frac{4}{\alpha+1}\left(\frac{2}{\alpha+1} + \frac{1}{\alpha+2}\right) + \frac{3}{(\alpha+2)^2}\right) + \frac{15}{(\alpha+2)^3}\Big)$$
$$\left. + \frac{8}{\alpha+1}\left(\frac{6}{\alpha+1}\left(\frac{4}{\alpha+1}\left(\frac{2}{\alpha+1} + \frac{1}{\alpha+2}\right) + \frac{3}{(\alpha+2)^2}\right) + \frac{15}{(\alpha+2)^3}\right) + \frac{105}{(\alpha+2)^4}\right.$$

$$a_{10,4}(\alpha) = \frac{9}{\alpha}\left(\frac{7}{\alpha}\left(\frac{5}{\alpha}\left(\frac{4}{\alpha+1} + \frac{3}{\alpha+2} + \frac{2}{\alpha+3} + \frac{1}{\alpha+4}\right) + \frac{6}{\alpha+1}\left(\frac{4}{\alpha+1} + \frac{3}{\alpha+2} + \frac{2}{\alpha+3} + \frac{1}{\alpha+4}\right)\right.\right.$$
$$+ \frac{5}{\alpha+2}\left(\frac{3}{\alpha+2} + \frac{2}{\alpha+3} + \frac{1}{\alpha+4}\right) + \frac{4}{\alpha+3}\left(\frac{2}{\alpha+3} + \frac{1}{\alpha+4}\right) + \frac{3}{(\alpha+4)^2}\Big)$$
$$+ \frac{8}{\alpha+1}\left(\frac{6}{\alpha+1}\left(\frac{4}{\alpha+1} + \frac{3}{\alpha+2} + \frac{2}{\alpha+3} + \frac{1}{\alpha+4}\right)\right.$$
$$+ \frac{5}{\alpha+2}\left(\frac{3}{\alpha+2} + \frac{2}{\alpha+3} + \frac{1}{\alpha+4}\right) + \frac{4}{\alpha+3}\left(\frac{2}{\alpha+3} + \frac{1}{\alpha+4}\right) + \frac{3}{(\alpha+4)^2}\Big)$$
$$+ \frac{7}{\alpha+2}\left(\frac{5}{\alpha+2}\left(\frac{3}{\alpha+2} + \frac{2}{\alpha+3} + \frac{1}{\alpha+4}\right) + \frac{4}{\alpha+3}\left(\frac{2}{\alpha+3} + \frac{1}{\alpha+4}\right) + \frac{3}{(\alpha+4)^2}\right)$$
$$\left. + \frac{6}{\alpha+3}\left(\frac{4}{\alpha+3}\left(\frac{2}{\alpha+3} + \frac{1}{\alpha+4}\right) + \frac{3}{(\alpha+4)^2}\right) + \frac{15}{(\alpha+4)^3}\right.$$

$$a_{10,6}(\alpha) = \frac{9}{\alpha}\left(\frac{7}{\alpha} + \frac{6}{\alpha+1} + \frac{5}{\alpha+2} + \frac{4}{\alpha+3} + \frac{3}{\alpha+4} + \frac{2}{\alpha+5} + \frac{1}{\alpha+6}\right)$$
$$+ \frac{8}{\alpha+1}\left(\frac{6}{\alpha+1} + \frac{5}{\alpha+2} + \frac{4}{\alpha+3} + \frac{3}{\alpha+4} + \frac{2}{\alpha+5} + \frac{1}{\alpha+6}\right)$$
$$+ \frac{7}{\alpha+2}\left(\frac{5}{\alpha+2} + \frac{4}{\alpha+3} + \frac{3}{\alpha+4} + \frac{2}{\alpha+5} + \frac{1}{\alpha+6}\right)$$
$$+ \frac{6}{\alpha+3}\left(\frac{4}{\alpha+3} + \frac{3}{\alpha+4} + \frac{2}{\alpha+5} + \frac{1}{\alpha+6}\right)$$
$$+ \frac{5}{\alpha+4}\left(\frac{3}{\alpha+4} + \frac{2}{\alpha+5} + \frac{1}{\alpha+6}\right) + \frac{4}{\alpha+5}\left(\frac{2}{\alpha+5} + \frac{1}{\alpha+6}\right) + \frac{3}{(\alpha+6)^2}$$

$$a_{10,8}(\alpha) = \frac{9}{\alpha} + \frac{8}{\alpha+1} + \frac{7}{\alpha+2} + \frac{6}{\alpha+3} + \frac{5}{\alpha+4} + \frac{4}{\alpha+5} + \frac{3}{\alpha+6} + \frac{2}{\alpha+7} + \frac{1}{\alpha+8}$$

$$a_{10,10}(\alpha) = 1$$

The case $\alpha = 1$ is of special interest. On substitution the following numerical values are derived:

$s = 0$: $a_{0,0}(1) = 1$

$s = 1$: $a_{1,1}(1) = 1$

$s = 2$: $a_{2,0}(1) = 1$, $a_{2,2}(1) = 1$

$s = 3$: $a_{3,1}(1) = 2\frac{1}{2}$, $a_{3,3}(1) = 1$

$s = 4$: $a_{4,0}(1) = 3$, $a_{4,2}(1) = 4\frac{1}{3}$, $a_{4,4}(1) = 1$

$s = 5$: $a_{5,1}(1) = 10\frac{3}{4}$, $a_{5,3}(1) = 6\frac{5}{12}$, $a_{5,5}(1) = 1$

$s = 6$: $a_{6,0}(1) = 15$, $a_{6,2}(1) = 24\frac{2}{3}$, $a_{6,4}(1) = 8\frac{7}{10}$, $a_{6,6}(1) = 1$

$s = 7$: $a_{7,1}(1) = 66\frac{2}{3}$, $a_{7,3}(1) = 45\frac{137}{144}$, $a_{7,5}(1) = 11\frac{3}{20}$, $a_{7,7}(1) = 1$

$s = 8$: $a_{8,0}(1) = 105$, $a_{8,2}(1) = 182\frac{2}{9}$, $a_{8,4}(1) = 75\frac{49}{75}$, $a_{8,6}(1) = 13\frac{26}{35}$, $a_{8,8}(1) = 1$

$s = 9$: $a_{9,1}(1) = 537\frac{9}{16}$, $a_{9,3}(1) = 396\frac{143}{192}$, $a_{9,5}(1) = 114\frac{283}{400}$, $a_{9,7}(1) = 16\frac{129}{280}$, $a_{9,9}(1) = 1$

$s = 10$: $a_{10,0}(1) = 945$, $a_{10,2}(1) = 1679\frac{14}{27}$, $a_{10,4}(1) = 749\frac{691}{900}$, $a_{10,6}(1) = 163\frac{2116}{2205}$, $a_{10,8}(1) = 19\frac{73}{252}$, $a_{10,10}(1) = 1$

We conclude this section by quoting a simple numerical example which will make the use of the formulae (23) sufficiently clear. The second of the two formulae gives, for $k = 2$,

$$\phi(4; \alpha; \beta, \gamma) = [a_{4,0}(\alpha) + a_{4,2}(\alpha)\,(\mathbf{PL})^2 + a_{4,4}(\alpha)\,(\mathbf{PL})^4]\,\phi(0; \alpha; \beta, \gamma).$$

To illustrate further, let $\beta + \gamma < 4$ (sample size $n < 5$). Then, from (20), the last term in the above expansion vanishes for all permissible values of α. Hence, on using (10) and (16),

$$\begin{aligned}
\phi(4; \alpha; \beta, \gamma) &= a_{4,0}(\alpha)\,\phi(0; \alpha; \beta, \gamma) + a_{4,2}(\alpha)\,\mathbf{P}^2\mathbf{L}^2\phi(0; \alpha; \beta, \gamma) \\
&= a_{4,0}(\alpha)\,\phi(0; \alpha; \beta, \gamma) + a_{4,2}(\alpha)\,\mathbf{P}^2\{\beta(\beta-1)\,\mathbf{B}^2 - 2\beta\gamma\mathbf{B}\boldsymbol{\Gamma} + \gamma(\gamma-1)\,\boldsymbol{\Gamma}^2\} \\
&\qquad \times \phi(0; \alpha; \beta, \gamma) \\
&= a_{4,0}(\alpha)\,\phi(0; \alpha; \beta, \gamma) + a_{4,2}(\alpha)\,\mathbf{P}^2\{\beta(\beta-1)\,\phi(0; \alpha; \beta-2, \gamma) \\
&\qquad - 2\beta\gamma\phi(0; \alpha; \beta-1, \gamma-1) + \gamma(\gamma-1)\,\phi(0; \alpha; \beta, \gamma-2)\} \\
&= a_{4,0}(\alpha)\,\phi(0; \alpha; \beta, \gamma) + a_{4,2}(\alpha)\,\frac{1}{\alpha(\alpha+1)} \\
&\qquad \times \{\beta(\beta-1)\,\phi(0; \alpha+2; \beta-2, \gamma) - 2\beta\gamma\phi(0; \alpha+2; \beta-1, \gamma-1) \\
&\qquad + \gamma(\gamma-1)\,\phi(0; \alpha+2; \beta, \gamma-2)\} \\
&= \phi(0; \alpha; \beta, \gamma) + \left(\frac{1}{\alpha+2} + \frac{2}{\alpha+1} + \frac{3}{\alpha}\right)\frac{1}{\alpha(\alpha+1)} \\
&\qquad \times \{\beta(\beta-1)\,\phi(0; \alpha+2; \beta-2, \gamma) - 2\beta\gamma\phi(0; \alpha+2; \beta-1, \gamma-1) \\
&\qquad + \gamma(\gamma-1)\,\phi(0; \alpha+2; \beta, \gamma-2)\}, \quad \beta + \gamma < 4.
\end{aligned}$$

In particular,

$$\phi(4;1;\beta,\gamma)=\phi(0;1;\beta,\gamma)+4\tfrac{1}{2}\times\tfrac{1}{2}\{\beta(\beta-1)\phi(0;3;\beta-2,\gamma)-2\beta\gamma\phi(0;3;\beta-1,\gamma-1)+\gamma(\gamma-1)\phi(0;3;\beta,\gamma-2)\},\quad \beta+\gamma<4.$$

2. THE GEOMETRICAL SIGNIFICANCE OF $\phi(0;\alpha;\beta,\gamma)$

The case of $\beta=0$, $\gamma=0$ is trivial. In this section $\beta+\gamma$ is a positive integer. We have

$$\begin{aligned}\phi(0;\alpha;\beta,\gamma)&=\int_{-\infty}^{\infty}x^{\alpha}p^{\beta}q^{\gamma}\,dx\\&=\int_{-\infty}^{\infty}\frac{1}{(2\pi)^{\frac{1}{2}\alpha}}e^{-\frac{1}{2}\alpha x^2}\prod_{i=1}^{\beta}\left(\int_{-\infty}^{x}\frac{1}{(2\pi)^{\frac{1}{2}}}e^{-\frac{1}{2}u_i^2}du_i\right)\prod_{j=\beta+1}^{\beta+\gamma}\left(\int_{x}^{\infty}\frac{1}{(2\pi)^{\frac{1}{2}}}e^{-\frac{1}{2}u_j}du_j\right)dx\\&=\frac{1}{(2\pi)^{\frac{1}{2}(\alpha+\beta+\gamma)}}\int_{u_1=-\infty}^{x}\int_{u_2=-\infty}^{x}\cdots\int_{u_\beta=-\infty}^{x}\\&\quad\times\int_{u_{\beta+1}=x}^{\infty}\int_{u_{\beta+2}=x}^{\infty}\cdots\int_{u_{\beta+\gamma}=x}^{\infty}\int_{x=-\infty}^{\infty}\exp\left[-\frac{1}{2}\left(\alpha x^2+\sum_1^{\beta+\gamma}u_i^2\right)\right]dx\prod_1^{\beta+\gamma}du_i.\end{aligned}\tag{32}$$

On applying the transformation

$$\left.\begin{aligned}y_i&=u_i-x\quad(i=1,2,\ldots,\beta+\gamma),\\x'&=x,\end{aligned}\right\}\tag{33}$$

$$\begin{aligned}\phi(0;\alpha;\beta,\gamma)=\frac{1}{(2\pi)^{\frac{1}{2}(\alpha+\beta+\gamma)}}&\int_{y_1=-\infty}^{0}\int_{y_2=-\infty}^{0}\cdots\int_{y_\beta=-\infty}^{0}\int_{y_{\beta+1}=0}^{\infty}\int_{y_{\beta+2}=0}^{\infty}\cdots\int_{y_{\beta+\gamma}=0}^{\infty}\int_{x'=-\infty}^{\infty}\\&\times\exp\left[-\tfrac{1}{2}(\alpha+\beta+\gamma)\left(x'+\sum_1^{\beta+\gamma}y_i\Big/(\alpha+\beta+\gamma)\right)^2-\frac{Q}{2(\alpha+\beta+\gamma)}\right]dx'\prod_1^{\beta+\gamma}dy_i,\end{aligned}\tag{34}$$

where Q is the positive-definite quadratic form

$$Q(y_1,y_2,\ldots,y_{\beta+\gamma})=(\alpha+\beta+\gamma-1)\Sigma y_i^2-\Sigma'\Sigma'y_iy_j,\tag{35}$$

the accented summations indicating that $j\neq i$. Integrating with respect to x',

$$\begin{aligned}\phi(0;\alpha;\beta,\gamma)=\frac{1}{(\alpha+\beta+\gamma)^{\frac{1}{2}}(2\pi)^{\frac{1}{2}(\alpha+\beta+\gamma-1)}}&\int_{y_1=-\infty}^{0}\cdots\int_{y_\beta=-\infty}^{0}\int_{y_{\beta+1}=0}^{\infty}\cdots\int_{y_{\beta+\gamma}=0}^{\infty}\\&\times\exp\left[-\frac{Q}{2(\alpha+\beta+\gamma)}\right]dy_1dy_2\ldots dy_{\beta+\gamma}.\end{aligned}\tag{36}$$

The right-hand member of the equation (36) may be interpreted as a quantity proportional† to the total probability-mass in one of the generalized quadrants of the multivariate normal distribution with dispersion matrix $\alpha+\beta+\gamma$ times the reciprocal of the matrix of Q. The characteristic numbers of the matrix of $Q/(\alpha+\beta+\gamma-1)$ are $(\alpha+\beta+\gamma)/(\alpha+\beta+\gamma-1)$ with multiplicity $\beta+\gamma-1$ and $\alpha/(\alpha+\beta+\gamma-1)$ with multiplicity 1. Hence, the orthogonal transformation

$$\left.\begin{aligned}y_1&=\sum_1^{\beta+\gamma}\xi_i/\sqrt{(\beta+\gamma)},\\y_s&=\sum_1^{\beta+\gamma}b_{si}\xi_i\quad(s=2,3,\ldots,\beta+\gamma),\end{aligned}\right\}\tag{37}$$

† The proportionality factor is $(2\pi)^{-\frac{1}{2}(\alpha-1)}\alpha^{-\frac{1}{2}}$.

reduces Q to a sum of squares,

$$Q(y_1, y_2, \ldots, y_{\beta+\gamma}) \equiv (\alpha+\beta+\gamma-1)\left(\frac{\alpha+\beta+\gamma}{\alpha+\beta+\gamma-1}\sum_1^{\beta+\gamma-1}\xi_i^2 + \frac{\alpha}{\alpha+\beta+\gamma-1}\xi_{\beta+\gamma}^2\right). \quad (38)$$

With the further scaling transformation

$$\left.\begin{aligned}\xi_i^* &= \left(\frac{\alpha+\beta+\gamma}{\alpha+\beta+\gamma-1}\right)^{\frac{1}{2}}\xi_i \quad (i = 1, 2, \ldots, \beta+\gamma-1),\\ \xi_{\beta+\gamma}^* &= \left(\frac{\alpha}{\alpha+\beta+\gamma-1}\right)^{\frac{1}{2}}\xi_{\beta+\gamma},\end{aligned}\right\} \quad (39)$$

$$Q(y_1, y_2, \ldots, y_{\beta+\gamma}) \equiv (\alpha+\beta+\gamma-1)\sum_1^{\beta+\gamma}\xi_i^{*2}. \quad (40)$$

The relation between the y_i and the ξ_i^* is provided by

$$\left.\begin{aligned}y_1 &= \frac{1}{(\beta+\gamma)^{\frac{1}{2}}}\left[\left(\frac{\alpha+\beta+\gamma-1}{\alpha+\beta+\gamma}\right)^{\frac{1}{2}}\sum_1^{\beta+\gamma-1}\xi_i^* + \left(\frac{\alpha+\beta+\gamma-1}{\alpha}\right)^{\frac{1}{2}}\xi_{\beta+\gamma}^*\right],\\ y_s &= \left(\frac{\alpha+\beta+\gamma-1}{\alpha+\beta+\gamma}\right)^{\frac{1}{2}}\sum_1^{\beta+\gamma-1}b_{si}\xi_i^* + \left(\frac{\alpha+\beta+\gamma-1}{\alpha}\right)^{\frac{1}{2}}b_{s,\beta+\gamma}\xi_{\beta+\gamma}^* \quad (s = 2, 3, \ldots, \beta+\gamma).\end{aligned}\right\} \quad (41)$$

Defining

$$\left.\begin{aligned}L_1 &= -\frac{1}{(\beta+\gamma)^{\frac{1}{2}}}\left[\left(\frac{\alpha+\beta+\gamma-1}{\alpha+\beta+\gamma}\right)^{\frac{1}{2}}\sum_1^{\beta+\gamma-1}\xi_i^* + \left(\frac{\alpha+\beta+\gamma-1}{\alpha}\right)^{\frac{1}{2}}\xi_{\beta+\gamma}^*\right],\\ L_j &= -\left(\frac{\alpha+\beta+\gamma-1}{\alpha+\beta+\gamma}\right)^{\frac{1}{2}}\sum_1^{\beta+\gamma-1}b_{ji}\xi_i^* - \left(\frac{\alpha+\beta+\gamma-1}{\alpha}\right)^{\frac{1}{2}}b_{j,\beta+\gamma}\xi_{\beta+\gamma}^* \quad (j = 2, 3, \ldots, \beta),\\ L_k &= \left(\frac{\alpha+\beta+\gamma-1}{\alpha+\beta+\gamma}\right)^{\frac{1}{2}}\sum_1^{\beta+\gamma-1}b_{ki}\xi_i^* + \left(\frac{\alpha+\beta+\gamma-1}{\alpha}\right)^{\frac{1}{2}}b_{k,\beta+\gamma}\xi_{\beta+\gamma}^* \quad (k = \beta+1, \beta+2, \ldots, \beta+\gamma),\end{aligned}\right\} \quad (42)$$

the region of integration in the y-space, $y_i \leqslant 0\ (i = 1, 2, \ldots, \beta), y_i \geqslant 0\ (i = \beta+1, \beta+2, \ldots, \beta+\gamma)$ transforms into the following region R in the ξ^*-space, demarcated by the bounding planes $L_i = 0$: $L_i \geqslant 0\ (i = 1, 2, \ldots, \beta+\gamma)$.

Denote the plane angle *interior* to R and formed by the bounding planes L_i and L_j $(j \neq i)$ by (ij).

There is no loss of generality if we set

$$b_{j,\beta+\gamma} = (\beta+\gamma)^{-\frac{1}{2}} \quad (j = 2, 3, \ldots, \beta+\gamma). \quad (43)$$

On using (43) together with the orthogonality relationships

$$\sum_{s=1}^{\beta+\gamma} b_{js}^2 = 1,$$

$$\sum_{s=1}^{\beta+\gamma} b_{js} = 0,$$

we find

$$\cos(1j) = -\frac{\frac{1}{(\beta+\gamma)^{\frac{1}{2}}}\frac{\alpha+\beta+\gamma-1}{\alpha+\beta+\gamma}\sum_{s=1}^{\beta+\gamma-1}b_{js} + \frac{1}{(\beta+\gamma)^{\frac{1}{2}}}\frac{\alpha+\beta+\gamma-1}{\alpha}b_{j,\beta+\gamma}}{\left[(\beta+\gamma-1)\frac{\alpha+\beta+\gamma-1}{\alpha+\beta+\gamma}\frac{1}{\beta+\gamma} + \frac{\alpha+\beta+\gamma-1}{\alpha}\frac{1}{\beta+\gamma}\right]^{\frac{1}{2}}\left[\frac{\alpha+\beta+\gamma-1}{\alpha+\beta+\gamma}\sum_{s=1}^{\beta+\gamma-1}b_{js}^2 + \frac{\alpha+\beta+\gamma-1}{\alpha}b_{j,\beta+\gamma}^2\right]^{\frac{1}{2}}}$$

$$= -\frac{1}{1+\alpha} \quad (j = 2, 3, \ldots, \beta).$$

Similarly,
$$\cos(1k) = \frac{1}{1+\alpha} \quad (k = \beta+1, \beta+2, \ldots, \beta+\gamma).$$

In exactly the same way it may be shown that

$$\left.\begin{aligned} \cos(ij) &= -\frac{1}{1+\alpha}; \quad i,j \text{ together in either of the blocks } (1,2,\ldots,\beta), \\ &\qquad\qquad (\beta+1,\beta+2,\ldots,\beta+\gamma),\ j \neq i \\ &= \frac{1}{1+\alpha}; \quad i,j \text{ in different blocks } (1,2,\ldots,\beta),\ (\beta+1,\beta+2,\ldots,\beta+\gamma), \end{aligned}\right\} \tag{44}$$

on using the orthogonality relationships

$$\begin{aligned} \sum_{s=1}^{\beta+\gamma} b_{is} b_{js} &= 0 \quad (j \neq i) \\ &= 1 \quad (j = i). \end{aligned}$$

The Jacobian of the transformation relating the y_i to the ξ_i^* is

$$\left|\frac{\partial(y_1, y_2, \ldots, y_{\beta+\gamma})}{\partial(\xi_1^*, \xi_2^*, \ldots, \xi_{\beta+\gamma}^*)}\right| = \left(\frac{\alpha+\beta+\gamma-1}{\alpha+\beta+\gamma}\right)^{\frac{1}{2}(\beta+\gamma-1)} \left(\frac{\alpha+\beta+\gamma-1}{\alpha}\right)^{\frac{1}{2}}.$$

Hence

$$\begin{aligned} &\phi(0; \alpha; \beta, \gamma) \\ &= \frac{1}{(2\pi)^{\frac{1}{2}(\alpha+\beta+\gamma-1)} \alpha^{\frac{1}{2}}} \left(\frac{\alpha+\beta+\gamma-1}{\alpha+\beta+\gamma}\right)^{\frac{1}{2}(\beta+\gamma)} \int_R \cdots \int \exp\left[-\frac{\alpha+\beta+\gamma-1}{2(\alpha+\beta+\gamma)} \sum_1^{\beta+\gamma} \xi_i^{*2}\right] \prod_1^{\beta+\gamma} d\xi_i^*. \end{aligned} \tag{45}$$

Finally, apply the spherical polar transformation

$$\left.\begin{aligned} \xi_i^* &= r \sin\chi_1 \sin\chi_2 \ldots \sin\chi_{i-1} \cos\chi_i \quad (i = 1, 2, \ldots, \beta+\gamma-1), \\ \xi_{\beta+\gamma}^* &= r \sin\chi_1 \sin\chi_2 \ldots \sin\chi_{\beta+\gamma-1}. \end{aligned}\right\} \tag{46}$$

Then

$$\phi(0; \alpha; \beta, \gamma) = \frac{1}{(2\pi)^{\frac{1}{2}(\alpha+\beta+\gamma-1)} \alpha^{\frac{1}{2}}} \left(\frac{\alpha+\beta+\gamma-1}{\alpha+\beta+\gamma}\right)^{\frac{1}{2}(\beta+\gamma)} \int_R \cdots \int \exp\left[-\frac{\alpha+\beta+\gamma-1}{2(\alpha+\beta+\gamma)} r^2\right] r^{\beta+\gamma-1} dr\, d\omega, \tag{47}$$

where $d\omega$ represents an infinitesimal solid angle in a space of $\beta+\gamma$ dimensions:

$$d\omega = \sin^{\beta+\gamma-2}\chi_1 \sin^{\beta+\gamma-3}\chi_2 \ldots \sin^2\chi_{\beta+\gamma-3} \sin\chi_{\beta+\gamma-2} d\chi_1 d\chi_2 \ldots d\chi_{\beta+\gamma-1}. \tag{48}$$

Integrating with respect to r $(0 \leqslant r \leqslant \infty)$,

$$\phi(0; \alpha; \beta, \gamma) = \frac{\Gamma\left(\frac{\beta+\gamma}{2}\right)}{2^{\frac{1}{2}(\alpha+1)} \pi^{\frac{1}{2}(\alpha+\beta+\gamma-1)} \alpha^{\frac{1}{2}}} \int_S d\omega, \tag{49}$$

S being the region on the unit sphere $\sum_1^{\beta+\gamma} \xi_i^{*2} = 1$ which is contained within the region R, so that $\int_S d\omega$ is simply the area of that portion of the unit sphere.† Denote the *ratio* which this area bears to the area of the unit sphere by $V_{\beta,\beta+\gamma}\left(\cos^{-1}\frac{1}{1+\alpha}\right)$. Equation (49) may then be rewritten in the form

$$\phi(0; \alpha; \beta, \gamma) = \frac{1}{(2\pi)^{\frac{1}{2}(\alpha-1)} \alpha^{\frac{1}{2}}} V_{\beta,\beta+\gamma}\left(\cos^{-1}\frac{1}{1+\alpha}\right). \tag{50}$$

† S is a (hyper)spherical simplex. For a discussion on hyperspherical simplices and their properties see, for example, Sommerville (1929) and Coxeter (1948).

The following result consequential on (50) is noted. The moment-generating function of the square of the rth order statistic, counting from the lowest member, in a random sample of n items from a normal population is provided by

$$\begin{aligned}\Phi_{r|n}(\theta) &= \mathscr{E}(e^{\theta x^2}) \\ &= \mathscr{E}(z^{-2\theta})\,(2\pi)^{-\theta} \\ &= (2\pi)^{-\theta}\int_{-\infty}^{\infty} z^{-2\theta}\,\frac{n!}{(r-1)!\,(n-r)!}\,zp^{r-1}q^{n-r}\,dx \\ &= (2\pi)^{-\theta}\,\frac{n!}{(r-1)!\,(n-r)!}\,\phi(0;\,1-2\theta;\,r-1,n-r)(\theta<\tfrac{1}{2}) \\ &= \frac{n!}{(r-1)!\,(n-r)!}\,V_{r-1,n-1}\left(\cos^{-1}-\frac{1}{2-2\theta}\right)(1-2\theta)^{-\frac{1}{2}} \quad (n=1,2,\ldots;\,r=1,2,\ldots,n).\end{aligned} \tag{51}$$

The interesting property follows from (51) that the square of any order statistic in samples from a normal population is distributed as the sum of the two independent variates, one of which is a χ^2 on one degree of freedom and the other a variate which has a moment-generating function provided by the first factor on the right-hand side of (51). In particular,

$$\Phi_{r|2}(\theta) = (1-2\theta)^{-\frac{1}{2}} \quad (r=1,2), \tag{52}$$

i.e. the even moments of either the smaller or of the greater member in a random sample of two items from a normal population are identical with the even moments relating to a sample of one item. Furthermore,

$$\Phi_{1|3}(\theta) = \frac{3\cos^{-1}\left(-\dfrac{1}{2-2\theta}\right)}{2\pi}(1-2\theta)^{-\frac{1}{2}} = \Phi_{3|3}(\theta), \tag{53}$$

$$\Phi_{2|3}(\theta) = 3\left(1-\frac{\cos^{-1}\left(-\dfrac{1}{2-2\theta}\right)}{\pi}\right)(1-2\theta)^{-\frac{1}{2}}, \tag{54}$$

$$\Phi_{1|4}(\theta) = \left(\frac{3\cos^{-1}\left(-\dfrac{1}{2-2\theta}\right)}{\pi}-1\right)(1-2\theta)^{-\frac{1}{2}} = \Phi_{4|4}(\theta), \tag{55}$$

$$\Phi_{2|4}(\theta) = 3\left(1-\frac{\cos^{-1}\left(-\dfrac{1}{2-2\theta}\right)}{\pi}\right)(1-2\theta)^{-\frac{1}{2}} = \Phi_{3|4}(\theta). \tag{56}$$

The results (52)–(56) follow from (51) on noting that

$$V_{0,1}(\theta) = \tfrac{1}{2} = V_{1,1}(\theta), \tag{57}$$

$$V_{0,2}(\theta) = \frac{\theta}{2\pi} = V_{2,2}(\theta), \tag{58}$$

$$V_{1,2}(\theta) = \frac{1}{2}\left(1-\frac{\theta}{\pi}\right), \tag{59}$$

$$V_{0,3}(\theta) = \frac{1}{4}\left(\frac{3\theta}{\pi}-1\right) = V_{3,3}(\theta), \tag{60}$$

$$V_{1,3}(\theta) = \frac{1}{4}\left(1-\frac{\theta}{\pi}\right) = V_{2,3}(\theta). \tag{61}$$

Equation (51) is also to be regarded as valid for $n = 1$ provided $V_{0,0}(\theta)$ is interpreted as 1.

The equations (57)–(61) express the fact that the surface contents of simplices constructed on 'spheres' which are immersed in one, two and three dimensions are, respectively, a point, an arc of a circle and a spherical triangle, the area of the latter being provided by the spherical excess. For dimensionality greater than three (spherical tetrahedra, spherical pentahedra, etc.) the areas can no longer be expressed in terms of elementary functions (§ 3). It will appear subsequently ((65) and (66)) that the *odd* moments of order statistics in samples of 2, 3 and 4 items, drawn from a normal population, can also be expressed in terms of elementary trigonometric quantities. This throws some light on the fact that some at any rate of the first few moments in samples of size $n \leqslant 4$ from normal populations have hitherto been available (see, for example, Jones (1948)) but not those relating to samples of size $n > 4$.

We note also that the equation (51) may be used to derive the even moments of the quantities, and, in particular, *all* the moments of the median (the odd moments of the latter being zero) in samples drawn from normal populations.

Reverting now to the question of the general moments, we substitute in (23) the expression given for $\phi(0; \alpha; \beta, \gamma)$ in (50), obtaining

$$\phi(2k+1; \alpha; \beta, \gamma) = \sum_{i=0}^{k} a_{2k+1,\,2i+1}(\alpha)\, \frac{1}{(2\pi)^{\frac{1}{2}(2i+\alpha)}\,(2i+\alpha+1)^{\frac{1}{2}}\,\alpha^{-(2i+1)}}\, \mathrm{L}^{2i+1} V_{\beta,\beta+\gamma}\left(\cos^{-1} - \frac{1}{2i+\alpha+2}\right) \quad (k = 0, 1, 2, \ldots), \tag{62}$$

$$\phi(2k; \alpha; \beta, \gamma) = \sum_{i=0}^{k} a_{2k,\,2i}(\alpha)\, \frac{1}{(2\pi)^{\frac{1}{2}(2i+\alpha-1)}\,(2i+\alpha)^{\frac{1}{2}}\,\alpha^{-(2i)}}\, \mathrm{L}^{2i} V_{\beta,\beta+\gamma}\left(\cos^{-1} - \frac{1}{2i+\alpha+1}\right) \quad (k = 0, 1, 2, \ldots). \tag{63}$$

On substituting further $\alpha = 1$, the moments of the rth order statistic, in a sample of size n from a normal distribution, defined by

$${}_s\mu'_{r\,|\,n} = n\binom{n-1}{r-1} \int_{-\infty}^{\infty} x^s z\, p^{r-1} q^{n-r}\, dx \quad (s = 1, 2, \ldots), \tag{64}$$

are provided by the following double series:

$${}_{2k+1}\mu'_{r\,|\,n} = n\binom{n-1}{r-1} \sum_{i=0}^{k} a_{2k+1,\,2i+1}(1)\, \frac{1}{(2\pi)^{i+\frac{1}{2}}\sqrt{(2i+2)\,(2i+1)!}}\, \mathrm{L}^{2i+1} V_{r-1,\,n-1}\left(\cos^{-1} - \frac{1}{2i+3}\right) \quad (k = 0, 1, 2, \ldots), \tag{65}$$

$${}_{2k}\mu'_{r\,|\,n} = n\binom{n-1}{r-1} \sum_{i=0}^{k} a_{2k,\,2i}(1)\, \frac{1}{(2\pi)^{i}\sqrt{(2i+1)\,(2i)!}}\, \mathrm{L}^{2i} V_{r-1,\,n-1}\left(\cos^{-1} - \frac{1}{2i+2}\right) \quad (k = 0, 1, 2, \ldots), \tag{66}$$

having put

$$\beta \equiv r-1, \quad \gamma \equiv n-r. \tag{67}$$

In (65) and in (66) it is to be remembered, as in (20), that

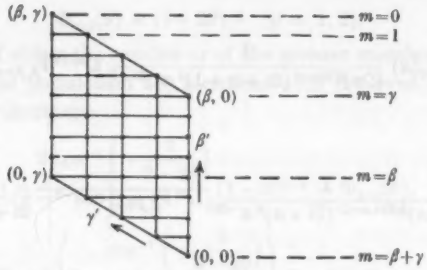
$$\mathrm{L}^m V_{\beta,\beta+\gamma}(\theta) \equiv 0 \quad (m = \beta+\gamma+1, \beta+\gamma+2, \ldots).$$

The character of the double series in (65) and in (66) may be represented graphically. For

$$\mathbf{L}^m V_{\beta,\beta+\gamma}(\theta) = \sum_{j=0}^{m} (-)^j \binom{m}{j} \beta^{(m-j)} \gamma^{(j)} V_{\beta+j-m,\beta+\gamma-m}(\theta), \tag{68}$$

and therefore the typical term on the right-hand side of (68) does not vanish, if, and only if, $m-\beta \leqslant j \leqslant \gamma$.

Let (β', γ') represent the co-ordinates of a point relative to two axes in a plane. Associate with each such point the function $V_{\beta',\beta'+\gamma'}$. Consider the non-negative integral points contained within and on the sides of the parallelogram with vertices $(0,0), (\beta,0), (0,\gamma), (\beta,\gamma)$. Then the value of the series in (68) is obtained by summation along the non-negative integral points on that portion of the line $\beta'+\gamma' = \beta+\gamma-m$ which lies within and on the sides of the parallelogram. Correspondingly, the values of the series in (65) and in (66) are obtained by summation along a set of such alternately spaced lines. For the extreme members, $\gamma = 0$, and the parallelogram shrinks into the line with end-points $(0,0)$ and $(\beta,0)$. Summation is then effected along the appropriate non-negative integral points on this line segment (see equations (100) and (101)).



Illustrating the character, and possible mode of summation, of the series in equations (65), (66) and (68).

It is worth noting from (50) that

$$V_{\beta,\beta+\gamma}(\tfrac{2}{3}\pi) = \frac{1}{\beta+\gamma+1} \frac{\beta!\gamma!}{(\beta+\gamma)!}, \tag{69}$$

and in particular,

$$V_{\beta,\beta}(\tfrac{2}{3}\pi) = \frac{1}{\beta+1}. \tag{70}$$

The relationship (70) is readily interpreted. For $V_{\beta,\beta}(\frac{2}{3}\pi)$ represents the relative content of a *regular* hyperspherical simplex, with primary bounding angles equal to $\frac{2}{3}\pi$, constructed on a sphere immersed in a space of β dimensions. Inscribe a regular generalized tetrahedron (with the angle between any two faces equal to $\cos^{-1} 1/\beta$) within the sphere. This generalized tetrahedron may be divided into $\beta+1$ equal simplices by a set of diametral planes which are inclined to each other at an angle of $\frac{2}{3}\pi$, irrespective of the dimensionality β. Hence this set of planes divides also the surface of the sphere into $\beta+1$ equal and regular hyperspherical simplices.

For purposes of illustration we end this section by considering the trivial yet illuminating case where $n = 4, r = 3$ ($\beta = 2, \gamma = 1$) and $k = 1$ in (66). This will show how the exact values of *all* the moments of order statistics in samples of size $n \leqslant 4$, some of which have been

obtained by Jones (1948), may be derived immediately from the fundamental formulae (65) and (66) once the values of the coefficients $a(1)$ are known. At the same time it will illustrate the use of these formulae in more difficult cases.

On substitution in (66),

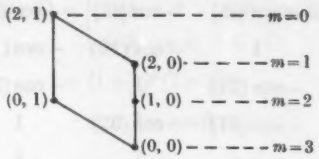
$$
\begin{aligned}
{}_2\mu'_{3|4} &= 4\binom{3}{2}\left\{a_{2,0}(1)V_{2,3}(\cos^{-1}-\tfrac{1}{4})+a_{2,2}(1)\frac{1}{2\pi\sqrt{3}\,2!}\mathrm{L}^2V_{2,3}(\cos^{-1}-\tfrac{1}{4})\right\} \\
&= 4\binom{3}{2}\left\{1\frac{1}{4}\frac{2!\,1!}{3!}+1\frac{1}{4\pi\sqrt{3}}[2.1.V_{0,1}(\cos^{-1}-\tfrac{1}{4})-2.2.1.V_{1,1}(\cos^{-1}-\tfrac{1}{4})+1.0.V_{2,1}(\cos^{-1}-\tfrac{1}{4})]\right\},
\end{aligned}
$$

having used equation (69). One would expect to interpret $V_{2,1}(\theta)$ as being identically zero (more generally, $V_{h,k}(\theta)\equiv 0$, $h>k$). However, such an interpretation is not strictly needed, since $V_{2,1}(\cos^{-1}-\frac{1}{4})$ is preceded in the above expansion by a zero coefficient (more generally, $V_{h,k}(\theta)$ will always be preceded in the series expansion for the moments by a zero coefficient whenever $h>k$). Using the values for $V_{0,1}(\cos^{-1}-\frac{1}{4})$ and $V_{1,1}(\cos^{-1}-\frac{1}{4})$ as given in equation (57), we obtain

$$
{}_2\mu'_{3|4} = 1-\sqrt{3}/\pi,
$$

in agreement with Jones's result.

The terms needed in the expansion can be obtained more expeditiously from the parallelogram diagram, the coefficients being derived from the values of the $a(1)$'s and by the use of (68). The parallelogram appropriate to the case $\beta=2$, $\gamma=1$ is depicted below:



The points along which summation is effected are (2, 1) ($m=0$), and (0, 1) and (1, 0) ($m=2$). The corresponding V's are $V_{2,3}$ and $V_{0\,1}$ and $V_{1,1}$ respectively, as in the above expansion for ${}_2\mu'_{3|4}$.

3. The content, $V_{\beta,\beta+\gamma}(\theta)$, of a type of hyperspherical simplex

The problem of determining the area of the general hyperspherical simplex is one of great complexity, even when the number of dimensions is as low as 4 (Hoppe, 1882; Richmond, 1903; Coxeter, 1935), and little progress appears to have been made in this direction. However, if we confine ourselves to a *singly-infinite* class of hyperspherical simplices, the problem becomes manageable although new types of transcendental functions are needed in the solution. §§ 1 and 2 indicate that for the full development of the theory of order statistics in samples from normal populations, the investigation of the properties of such a class of simplices is essential.

The starting point for this investigation is Schläfli's (1950) differential recurrence relation for the area $S_{n'}$ of the general simplex constructed on the surface of a sphere in n' dimensions:

$$
dS_{n'} = \frac{1}{n'-2}\{S(\overline{12})\,d(12)+S(\overline{13})\,d(13)+\ldots+\dot{S}[\overline{(n'-1)\,n'}]\,d[(n'-1)\,n']\} \quad (n'>2), \tag{71}
$$

in which $(\overline{ij})$ $(j \neq i)$, denotes the edge, of dimensionality two less than that of the original simplex, formed by the ith and jth flats and $S(\overline{ij})$ its content (e.g. for $n' = 5$, $S(\overline{ij})$ is the area of a spherical triangle). More generally, denote the edge of the simplex, of dimensionality m less than that of the original simplex, formed jointly by the first, second, ..., mth flats, by $(\overline{123\ldots m})$, and a typical primary angle of this simplex by $(\overline{123\ldots m}\,ik)$. These angles may be expressed in terms of the determinants of certain bordered matrices with the basic primary angles as elements.

In fact (Schläfi, 1950, p. 62)

$$\cos(\overline{12\ldots m}\,ik) = -\frac{\Delta\binom{i}{k}123\ldots m\Big)}{\{\Delta(i\,123\ldots m)\,\Delta(k\,123\ldots m)\}^{\frac{1}{2}}}, \tag{72}$$

with

$$\Delta\binom{i}{k}123\ldots m\Big) = \begin{vmatrix} -\cos(ik) & -\cos(i1) & -\cos(i2) & -\cos(i3) & \ldots & -\cos(im) \\ -\cos(1k) & 1 & -\cos(12) & -\cos(13) & \ldots & -\cos(1m) \\ -\cos(2k) & -\cos(21) & 1 & -\cos(23) & \ldots & -\cos(2m) \\ -\cos(3k) & -\cos(31) & -\cos(32) & 1 & \ldots & -\cos(3m) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\cos(mk) & -\cos(m1) & -\cos(m2) & -\cos(m3) & \ldots & 1 \end{vmatrix}, \tag{73}$$

$$\Delta(i\,123\ldots m) = \begin{vmatrix} 1 & -\cos(i1) & -\cos(i2) & -\cos(i3) & \ldots & -\cos(im) \\ -\cos(1i) & 1 & -\cos(12) & -\cos(13) & \ldots & -\cos(1m) \\ -\cos(2i) & -\cos(21) & 1 & -\cos(23) & \ldots & -\cos(2m) \\ -\cos(3i) & -\cos(31) & -\cos(32) & 1 & \ldots & -\cos(3m) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\cos(mi) & -\cos(m1) & -\cos(m2) & -\cos(m3) & \ldots & 1 \end{vmatrix}. \tag{74}$$

In terms of *relative* areas (71) may be written

$$dV_{n'} = \frac{1}{2\pi}\sum_{j>i}\sum V_{n'-2}(\overline{ij})\,d(ij) \quad (n' = 2, 3, \ldots), \tag{71'}$$

with an obvious slight change in notation. In this standardized form the relationship holds also for $n' = 2$ provided $V_0(\theta)$ is as above interpreted as being 1.

The simplex under investigation is one constructed on a sphere immersed in a space of $\beta+\gamma$-dimensionality, and one demarcated by $\beta+\gamma$ $\beta+\gamma+1$-flats which may be divided into two sets $(1, 2, 3, \ldots, \beta)$ and $(\beta+1, \beta+2, \ldots, \beta+\gamma)$ in such a way that

$$\begin{aligned} (ij) &= \theta, \qquad i, j \text{ in same set} \\ &= \pi - \theta, \quad i, j \text{ in opposite sets} \quad (\tfrac{1}{2}\pi \leqslant \theta \leqslant \pi). \end{aligned} \tag{75}$$

From (72), (73) and (74),

$$\cos(\overline{ij}\,\sigma\tau) = \frac{-\Delta\binom{\sigma}{\tau}ij\Big)}{\{\Delta(\sigma ij)\,\Delta(\tau ij)\}^{\frac{1}{2}}}, \tag{76}$$

with

$$\Delta\begin{pmatrix}\sigma\\ \tau\end{pmatrix}ij = \begin{vmatrix} -\cos(\sigma\tau) & -\cos(\sigma i) & -\cos(\sigma j) \\ -\cos(i\tau) & 1 & -\cos(ij) \\ -\cos(j\tau) & -\cos(ji) & 1 \end{vmatrix}, \tag{77}$$

$$\Delta(\sigma ij) = \begin{vmatrix} 1 & -\cos(\sigma i) & -\cos(\sigma j) \\ -\cos(i\sigma) & 1 & -\cos(ij) \\ -\cos(j\sigma) & -\cos(ji) & 1 \end{vmatrix}, \tag{78}$$

$$\Delta(\tau ij) = \begin{vmatrix} 1 & -\cos(\tau i) & -\cos(\tau j) \\ -\cos(i\tau) & 1 & -\cos(ij) \\ -\cos(j\tau) & -\cos(ji) & 1 \end{vmatrix}. \tag{78'}$$

There are sixteen possibilities to consider, for i, j, σ and τ may each be in either of the two sets. A typical example, appropriate to the case wherein i, j, σ and τ are all in the first (or all in the second) set, gives

$$\Delta\begin{pmatrix}\sigma\\ \tau\end{pmatrix}ij = \begin{vmatrix} -\lambda & -\lambda & -\lambda \\ -\lambda & 1 & -\lambda \\ -\lambda & -\lambda & 1 \end{vmatrix}$$

$$= -\lambda(1+\lambda)^2,$$

$$\Delta(\sigma ij) = \begin{vmatrix} 1 & -\lambda & -\lambda \\ -\lambda & 1 & -\lambda \\ -\lambda & -\lambda & 1 \end{vmatrix} = \Delta(\tau ij)$$

$$= (1+\lambda)^2(1-2\lambda),$$

where

$$\lambda \equiv \cos\theta. \tag{79}$$

Hence, in this case,

$$\cos(\overline{ij}\,\sigma\tau) = \frac{\lambda}{1-2\lambda}.$$

Similar examination of the third-order determinants involved reveals that

$$\cos(\overline{ij}\,\sigma\tau) = \frac{\lambda}{1-2\lambda} \equiv \cos(\phi(\theta)); \qquad \sigma, \tau \text{ in same set, all } i, j, j \neq i, \tau \neq \sigma$$

$$= \frac{-\lambda}{1-2\lambda} \equiv \cos(\pi - \phi(\theta)); \quad \sigma, \tau \text{ in opposite sets, all } i, j, j \neq i. \tag{80}$$

Consider $(\overline{ij})$, where i and j belong to the first set $(1, 2, \ldots, \beta)$ with $j \neq i$. This is a skew simplex of the type already discussed with dimensionality reduced by two. It may be regarded as demarcated by $\beta+\gamma-2$ $\beta+\gamma-3$-flats which divide into two groups $(1, 2, \ldots, i-1, i+1, \ldots, j-1, j+1, \ldots, \beta)$, $(\beta+1, \beta+2, \ldots, \beta+\gamma)$ such that the angle between any two distinct flats in the same group is $\phi(\theta)$ and that between any two flats in distinct groups is $\pi-\phi(\theta)$. Here also $d(ij) = d\theta$. Consider now $(\overline{ij})$, where i and j belong to the second set $(\beta+1, \beta+2, \ldots, \beta+\gamma)$ with $j \neq i$. This is also a skew simplex of the two already discussed with dimensionality reduced by two. It may be regarded as demarcated by $\beta+\gamma-2$ $\beta+\gamma-3$-flats which divide into two groups $(1, 2, \ldots, \beta)$, $(\beta+1, \ldots, i-1, i+1, \ldots, j-1, j+1, \ldots, \beta+\gamma)$, such that the angle between any two distinct flats in the same group is $\phi(\theta)$ and that between any two flats in different groups is $\pi-\phi(\theta)$. Here, likewise, $d(ij) = d\theta$. Consider, finally, $(\overline{ij})$,

where i and j are in opposite sets. This is once again a skew simplex of the type already discussed with dimensionality reduced by two. It may be regarded as demarcated by $\beta+\gamma-2$ $\beta+\gamma-3$-flats which divide into two groups $(1, 2, \ldots, i-1, i+1, \ldots, \beta)$, $(\beta+1, \beta+2, \ldots, j-1, j+1, \ldots, \beta+\gamma)$ such that the angle between any two distinct flats in the same group is $\phi(\theta)$ and that between any two flats in different groups is $\pi-\phi(\theta)$. Here, $d(ij) = -d\theta$.

Combining these results and applying them in (71′),

$$dV_{\beta,\beta+\gamma}(\theta) = \frac{1}{2\pi}\left\{\binom{\beta}{2} V_{\beta-2,\beta+\gamma-2}(\phi(\theta)) - \beta\gamma V_{\beta-1,\beta+\gamma-2}(\phi(\theta)) + \binom{\gamma}{2} V_{\beta,\beta+\gamma-2}(\phi(\theta))\right\} d\theta. \quad (81)$$

Hence

$$V_{\beta,\beta+\gamma}(\theta) = \frac{1}{2^{\beta+\gamma}} + \frac{1}{2\pi}\int_{\frac{1}{2}\pi}^{\theta}\left\{\binom{\beta}{2} V_{\beta-2,\beta+\gamma-2}(\phi(\theta')) - \beta\gamma V_{\beta-1,\beta+\gamma-2}(\phi(\theta')) + \binom{\gamma}{2} V_{\beta,\beta+\gamma-2}(\phi(\theta'))\right\} d\theta'$$

$$(\beta, \gamma = 0, 1, 2, \ldots). \quad (82)$$

Symbolically,

$$V_{\beta,\beta+\gamma}(\theta) = \frac{1}{2^{\beta+\gamma}} + \frac{1}{4\pi}\int_{\frac{1}{2}\pi}^{\theta} \mathbf{L}^2(V_{\beta,\beta+\gamma}(\phi(\theta')))\, d\theta'$$

$$= \frac{1}{2^{\beta+\gamma}} + \frac{1}{4\pi}\mathbf{L}^2\int_{\frac{1}{2}\pi}^{\theta} V_{\beta,\beta+\gamma}(\phi(\theta'))\, d\theta' \quad (\beta, \gamma = 0, 1, 2, \ldots). \quad (83)$$

Equation (83) may be used to express $V_{\beta,\beta+\gamma}(\theta)$ in terms of a set of auxiliary functions defined by

$$\psi_0(\theta) = 1,$$

$$\psi_k(\theta) = \frac{1}{\pi}\int_{\frac{1}{2}\pi}^{\theta} \psi_{k-1}(\phi(\theta'))\, d\theta' \quad (\tfrac{1}{2}\pi \leqslant \theta \leqslant \pi) \quad (k = 1, 2, \ldots). \quad (84)$$

In fact, by successive application of (83),

$$V_{\beta,\beta+\gamma}(\theta) = \frac{1}{2^{\beta+\gamma}} + \frac{1}{4\pi}\mathbf{L}^2\int_{\frac{1}{2}\pi}^{\theta}\left[\frac{1}{2^{\beta+\gamma}} + \frac{1}{4\pi}\mathbf{L}^2\int_{\frac{1}{2}\pi}^{\phi(\theta_1)} V_{\beta,\beta+\gamma}(\phi(\theta_2))\, d\theta_2\right] d\theta_1$$

$$= \frac{1}{2^{\beta+\gamma}} + \frac{1}{2^{\beta+\gamma}}\mathbf{L}^2\psi_1(\theta) + \frac{1}{(4\pi)^2}\int_{\frac{1}{2}\pi}^{\theta}\int_{\frac{1}{2}\pi}^{\phi(\theta_1)} \mathbf{L}^4 V_{\beta,\beta+\gamma}(\phi(\theta_2))\, d\theta_2 d\theta_1.$$

More generally,

$$V_{\beta,\beta+\gamma}(\theta)$$

$$= \frac{1}{2^{\beta+\gamma}}\sum_{k=0}^{s}\mathbf{L}^{2k}\psi_k(\theta) + \frac{1}{(4\pi)^{s+1}}\int_{\frac{1}{2}\pi}^{\theta}\int_{\frac{1}{2}\pi}^{\phi(\theta_1)}\int_{\frac{1}{2}\pi}^{\phi(\theta_2)}\cdots\int_{\frac{1}{2}\pi}^{\phi(\theta_s)} \mathbf{L}^{2s+2} V_{\beta,\beta+\gamma}(\phi(\theta_{s+1}))\, d\theta_{s+1} d\theta_s \ldots d\theta_2 d\theta_1. \quad (85)$$

There are two possibilities

(i) $\beta+\gamma$ even.

Setting $s = \frac{1}{2}(\beta+\gamma)-1$,

$$V_{\beta,\beta+\gamma}(\theta)$$

$$= \frac{1}{2^{\beta+\gamma}}\sum_{k=0}^{\frac{1}{2}(\beta+\gamma)-1}\mathbf{L}^{2k}\psi_k(\theta) + \frac{1}{(4\pi)^{\frac{1}{2}(\beta+\gamma)}}\int_{\frac{1}{2}\pi}^{\theta}\int_{\frac{1}{2}\pi}^{\phi(\theta_1)}\cdots\int_{\frac{1}{2}\pi}^{\phi(\theta_{\frac{1}{2}(\beta+\gamma)-1})} \mathbf{L}^{\beta+\gamma} V_{\beta,\beta+\gamma}(\phi(\theta_{\frac{1}{2}(\beta+\gamma)}))\, d\theta_{\frac{1}{2}(\beta+\gamma)} \ldots d\theta_1$$

$$= \frac{1}{2^{\beta+\gamma}}\sum_{k=0}^{\frac{1}{2}(\beta+\gamma)-1}\mathbf{L}^{2k}\psi_k(\theta) + \frac{1}{2^{\beta+\gamma}}(-)^{\beta+\gamma}(\beta+\gamma)!\,\psi_{\frac{1}{2}(\beta+\gamma)}(\theta), \quad (86)$$

since

$$\mathbf{L}^{\beta+\gamma}(V_{\beta,\beta+\gamma}(\theta)) \equiv (-)^{\gamma}(\beta+\gamma)!. \quad (87)$$

Again,
$$L^{\beta+\gamma}.1 = \sum_{j=0}^{\beta+\gamma}(-)^j\binom{\beta+\gamma}{j}\gamma^{(j)}\beta^{(\beta+\gamma-j)}$$
$$= (-)^\gamma\binom{\beta+\gamma}{\gamma}\gamma^{(\gamma)}\beta^{(\beta)}$$
$$= (-)^\gamma(\beta+\gamma)!. \tag{88}$$

Hence
$$V_{\beta,\beta+\gamma}(\theta) = \frac{1}{2^{\beta+\gamma}}\sum_{k=0}^{\frac{1}{2}(\beta+\gamma)}\mathbf{L}^{2k}\psi_k(\theta), \quad \beta+\gamma \text{ even}. \tag{89}$$

(ii) $\beta+\gamma$ odd.

Setting $s = \frac{1}{2}(\beta+\gamma-3)$,

$$V_{\beta,\beta+\gamma}(\theta) = \frac{1}{2^{\beta+\gamma}}\sum_{k=0}^{\frac{1}{2}(\beta+\gamma-3)}\mathbf{L}^{2k}\psi_k(\theta)$$
$$+\frac{1}{(4\pi)^{\frac{1}{2}(\beta+\gamma-1)}}\int_{\frac{1}{2}\pi}^{\theta}\int_{\frac{1}{2}\pi}^{\phi(\theta_1)}\cdots\int_{\frac{1}{2}\pi}^{\phi(\theta_{\frac{1}{2}(\beta+\gamma-3)})}\mathbf{L}^{\beta+\gamma-1}V_{\beta,\beta+\gamma}(\phi(\theta_{\frac{1}{2}(\beta+\gamma-1)}))\,d\theta_{\frac{1}{2}(\beta+\gamma-1)}\cdots d\theta_1$$
$$= \frac{1}{2^{\beta+\gamma}}\sum_{k=0}^{\frac{1}{2}(\beta+\gamma-3)}\mathbf{L}^{2k}\psi_k(\theta) + \frac{1}{2^{\frac{1}{2}(\beta+\gamma-1)}}\tfrac{1}{2}(\gamma-\beta)(\beta+\gamma-1)!\,\psi_{\frac{1}{2}(\beta+\gamma-1)}(\theta), \tag{90}$$

since
$$\mathbf{L}^{\beta+\gamma-1}(V_{\beta,\beta+\gamma}(\theta)) = \sum_{j=0}^{\beta+\gamma-1}(-)^j\binom{\beta+\gamma-1}{j}\gamma^{(j)}\beta^{(\beta+\gamma-j-1)}V_{j-\gamma+1,1}(\theta)$$
$$= (-)^{\gamma-1}\binom{\beta+\gamma-1}{\gamma-1}\beta^{(\beta)}\gamma^{(\gamma-1)}V_{0,1}(\theta) + (-)^\gamma\binom{\beta+\gamma-1}{\gamma}\beta^{(\beta-1)}\gamma^{(\gamma)}V_{1,1}(\theta)$$
$$\equiv -\tfrac{1}{2}(\beta-\gamma)(\beta+\gamma-1)!. \tag{91}$$

Again,
$$L^{\beta+\gamma-1}.1 = \sum_{j=0}^{\beta+\gamma-1}(-)^j\binom{\beta+\gamma-1}{j}\gamma^{(j)}\beta^{(\beta+\gamma-j-1)}$$
$$= -(\beta-\gamma)(\beta+\gamma-1)!. \tag{92}$$

Hence,
$$V_{\beta,\beta+\gamma}(\theta) = \frac{1}{2^{\beta+\gamma}}\sum_{k=0}^{\frac{1}{2}(\beta+\gamma-1)}\mathbf{L}^{2k}\psi_k(\theta), \quad \beta+\gamma \text{ odd}. \tag{93}$$

Combining (89) and (93),
$$V_{\beta,\beta+\gamma}(\theta) = \frac{1}{2^{\beta+\gamma}}\sum_{k=0}^{[\frac{1}{2}(\beta+\gamma)]}\mathbf{L}^{2k}\psi_k(\theta),\dagger \tag{94}$$

where $[m]$ denotes, as usual, the integral part of m.

The character of the double series in the right-hand member of (94) is similar to that of the series in (65) and (66), and the parallelogram diagram may be used to depict this character, as well as to effect summation. The values of $V_{\beta,\beta+\gamma}(\theta)$, and in particular the values of $V_{\beta,\beta+\gamma}\left(\cos^{-1}-\frac{1}{m}\right)$ $(m = 2, 3, 4, \ldots)$ needed for the specification of the moments in (65) and in (66), may be studied from the point of view of the auxiliary functions $\psi_k(\theta)$ and the special values $\psi_k\left(\cos^{-1}-\frac{1}{m}\right)$ $(m = 3, 4, 5, \ldots)$, or, alternatively, from the point of view of the chain

† It is known (Schläfli, 1950, pp. 240–3) that the content of a simplex on the surface of a sphere in $2m+1$ dimensions is a linear function of the contents of the simplices on the surfaces of spheres in dimensions, 2, 4, 6, ..., $2m$. Equation (94) illustrates this property in a particular instance.

relationship (83). In either case, in view of the special nature of the arguments of the ψ_k's and of the $V_{\beta,\beta+\gamma}$, the following transformation is useful:

$$\theta' = \cos^{-1}\left(-\frac{1}{x'}\right) \quad (x' \geqslant 1). \tag{95}$$

Defining

$$\overline{\psi}_k(x) \equiv \psi_k(\theta), \tag{96}$$

$$\overline{V}_{\beta,\beta+\gamma}(x) \equiv V_{\beta,\beta+\gamma}(\theta), \tag{97}$$

the following differential-difference equations for the $\overline{\psi}$'s, and the $\overline{V}_{\beta,\beta+\gamma}$ are derived from (84) and (83):

$$\overline{\psi}_k(x) = \int_x^\infty \overline{\psi}_{k-1}(x'+2)\frac{dx'}{x'\sqrt{(x'^2-1)}} \quad (k = 1, 2, \ldots), \tag{98}$$

$$\overline{V}_{\beta,\beta+\gamma}(x) = \frac{1}{2^{\beta+\gamma}} + \frac{1}{4\pi}\mathrm{L}^2\int_x^\infty \overline{V}_{\beta,\beta+\gamma}(x'+2)\frac{dx'}{x'\sqrt{(x'^2-1)}} \quad (\beta, \gamma = 0, 1, 2, \ldots). \tag{99}$$

The case of the extreme members of the sample ($\gamma = 0$, $r = n$) is of special interest. Equations (65) and (66) reduce to

$${}_{2k+1}\mu'_{n\,|\,n} = n\sum_{i=0}^{k} a_{2k+1,2i+1}(1)\binom{n-1}{2i+1}\frac{1}{(2\pi)^{i+\frac{1}{2}}\sqrt{(2+2i)}}V_{n-2-2i,n-2-2i}\left(\cos^{-1}-\frac{1}{3+2i}\right)$$

$$(k = 0, 1, 2, \ldots), \tag{100}$$

$${}_{2k}\mu'_{n\,|\,n} = n\sum_{i=0}^{k} a_{2k,2i}(1)\binom{n-1}{2i}\frac{1}{(2\pi)^{i}\sqrt{(1+2i)}}V_{n-1-2i,n-1-2i}\left(\cos^{-1}-\frac{1}{2+2i}\right) \quad (k = 0, 1, 2, \ldots). \tag{101}$$

The simplices involved in (100) and in (101) are *regular* with primary bounding angles $\cos^{-1}(-\frac{1}{3})$, $\cos^{-1}(-\frac{1}{5})$, ..., and $\cos^{-1}(-\frac{1}{2})$, $\cos^{-1}(-\frac{1}{4})$, ..., respectively. In terms of the $\overline{V}$'s,

$${}_{2k+1}\mu'_{n\,|\,n} = n\sum_{i=0}^{k} a_{2k+1,2i+1}(1)\binom{n-1}{2i+1}\frac{1}{(2\pi)^{i+\frac{1}{2}}\sqrt{(2i+2)}}\overline{V}_{n-2-2i,n-2-2i}(3+2i) \quad (k = 0, 1, 2, \ldots), \tag{100'}$$

$${}_{2k}\mu'_{n\,|\,n} = n\sum_{i=0}^{k} a_{2k,2i}(1)\binom{n-1}{2i}\frac{1}{(2\pi)^{i}\sqrt{(2i+1)}}\overline{V}_{n-1-2i,n-1-2i}(2+2i) \quad (k = 0, 1, 2, \ldots). \tag{101'}$$

In the last four equations, the combinational coefficient $\binom{r}{s}$ is to be interpreted as zero if $s > r$.

Equation (83) reduces to

$$V_{\beta,\beta}(\theta) = \frac{1}{2^\beta} + \frac{\beta(\beta-1)}{4\pi}\int_{\frac{1}{3}\pi}^{\theta} V_{\beta-2,\beta-2}(\phi(\theta'))\,d\theta' \quad (\beta = 0, 1, 2, \ldots), \tag{102}$$

or, in terms of the $\overline{V}$'s,

$$\overline{V}_{\beta,\beta}(x) = \frac{1}{2^\beta} + \frac{\beta(\beta-1)}{4\pi}\int_x^\infty \overline{V}_{\beta-2,\beta-2}(x'+2)\frac{dx'}{x'\sqrt{(x'^2-1)}} \quad (\beta = 0, 1, 2, \ldots). \tag{102'}$$

Equation (94) reduces to

$$V_{\beta,\beta}(\theta) = \frac{1}{2^\beta}\sum_{k=0}^{[\frac{1}{2}\beta]} \beta^{(2k)}\psi_k(\theta) \quad (\beta = 0, 1, 2, \ldots). \tag{103}$$

It is easy to verify from (70) and (103) that

$$\psi_k\left(\frac{2\pi}{3}\right) = \frac{1}{(2k+1)!} = \overline{\psi}_k(2) \quad (k = 0, 1, 2, \ldots). \tag{104}$$

Also, from (103) (together with the fact that $V_{\beta,\beta}(\pi) = \frac{1}{2}$),

$$\psi_k(\pi) = \frac{1}{(2k)!} = \overline{\psi}_k(1) \quad (k = 0, 1, 2, \ldots). \tag{105}$$

Finally, equation (51) reduces to

$$\begin{aligned}\Phi_{n\,|\,n}(\theta) &= nV_{n-1,n-1}\left(\cos^{-1} - \frac{1}{2-2\theta}\right)(1-2\theta)^{-\frac{1}{2}} \\ &= n\overline{V}_{n-1,n-1}(2-2\theta)(1-2\theta)^{-\frac{1}{2}}.\end{aligned} \tag{106}$$

4. COMPUTATION OF THE CONTENTS OF REGULAR HYPERSPHERICAL SIMPLICES AND THE ASSOCIATED MOMENTS OF THE EXTREME ORDER STATISTICS IN SAMPLES FROM NORMAL POPULATIONS

A knowledge of the contents of regular hyperspherical simplices is likely to be useful elsewhere in statistical applications and is not without theoretical interest. Table 1 provides the values of $\overline{V}_{\beta,\beta}(x)$ for $x = 2(1)12$ and $\beta = 1(1)49$. These are the values which arise in (100′) and in (101′) in the specification of the first ten moments of the extreme members in a sample containing up to fifty items drawn from a normal population. The values of the $\overline{V}_{\beta,\beta}(x)$ have been computed from the recurrence formula (102′). The values of the moments of the extreme members, given to a greater degree of accuracy than has hitherto been found computationally convenient, are provided in Table 2. These have been derived from equations (100′) and (101′), using the values of the $a(1)$'s given at the end of § 1 and the values of the $\overline{V}_{\beta,\beta}(x)$ given in table 1. Table 3 gives the second, third and fourth moments about the mean of the extreme members in samples of all sizes up to and including $n = 50$, as well as the standard deviations, together with β_1 and $\beta_2 - 3$.

It is hoped subsequently to compute by electronic means the values of the $\overline{V}_{\beta,\beta+\gamma}(x)$ (relative contents of skew hyperspherical simplices with primary bounding angles $\cos^{-1}(-1/x)$ and $\pi - \cos^{-1}(-1/x)$) from the generalized recurrence formula (99), and to apply these values in the general formulae (65) and (66) for the derivation of at least the first four moments of order statistics which are not extreme.

I am greatly indebted to Prof. Sir Ronald Fisher, F.R.S., for his advice and encouragement.

Table 1. *Relative surface contents of regular hyperspherical simplices* $\bar{u}_n(x)$ *with dimensionality* n *and primary bounding angles* $\cos^{-1}(-1/x)$†

n	$\bar{u}_{n-1}(2)$	$\bar{u}_{n-3}(4)$	$\bar{u}_{n-5}(6)$	$\bar{u}_{n-7}(8)$	$\bar{u}_{n-9}(10)$	$\bar{u}_{n-11}(12)$
1	1·00000 00000	—	—	—	—	—
2	0·50000 00000	—	—	—	—	—
3	0·33333 33333	1·00000 0000	—	—	—	—
4	0·25000 00000	0·50000 0000	—	—	—	—
5	0·20000 00000	0·29021 5312	1·00000 0000	—	—	—
6	0·16666 66667	0·18532 2967	0·50000 0000	—	—	—
7	0·14285 71429	0·12647 9249	0·27665 0190	1·00000 000	—	—
8	0·12500 00000	0·09065 9844	0·16497 5284	0·50000 000	—	—
9	0·11111 11111	0·06748 2726	0·10422 4047	0·26994 654	1·00000 00	—
10	0·10000 00000	0·05175 6879	0·06893 46456	0·15491 982	0·50000 00	—
11	0·09090 90909	0·04067 3686	0·04732 96970	0·09344 505	0·26594 21	1·00000 0
12	0·08333 33333	0·03261 5725	0·03352 05107	0·05874 626	0·14891 32	0·50000 0
13	0·07692 30769	0·02660 3196	0·02437 08312	0·03824 650	0·08708 71	0·26327 8
14	0·07142 85714	0·02201 7121	0·01812 06201	0·02565 733	0·05286 14	0·14491 7
15	0·06666 66667	0·01845 2415	0·01373 76418	0·01766 4546	0·03313 989	0·08289 4
16	0·06250 00000	0·01563 5828	0·01059 32558	0·01244 0956	0·02137 108	0·04904 0
17	0·05882 35294	0·01337 8162	0·00829 19269	0·00893 9452	0·01412 973	0·02988 9
18	0·05555 55556	0·01154 5394	0·00657 75941	0·00653 8976	0·00955 140	0·01870 1
19	0·05263 15789	0·01004 0653	0·00528 02708	0·00486 0042	0·00658 678	0·01198 4
20	0·05000 00000	0·00879 26628	0·00428 45634	0·00366 44887	0·00462 3852	0·00783 97
21	0·04761 90476	0·00774 81250	0·00351 05571	0·00279 92407	0·00329 9116	0·00523 58
22	0·04545 45455	0·00686 66186	0·00290 19034	0·00216 37714	0·00238 8923	0·00355 63
23	0·04347 82608	0·00611 70792	0·00241 82187	0·00169 07679	0·00175 3346	0·00245 48
24	0·04166 66667	0·00547 53518	0·00203 01296	0·00133 43527	0·00130 2901	0·00171 96
25	0·04000 00000	0·00492 24500	0·00171 59797	0·00106 27478	0·00097 9287	0·00122 121
26	0·03846 15384	0·00444 33011	0·00145 96035	0·00085 36129	0·00074 3854	0·00087 824
27	0·03703 70370	0·00402 58300	0·00124 87951	0·00069 10262	0·00057 0572	0·00063 900
28	0·03571 42857	0·00366 02813	0·00107 42412	0·00056 34963	0·00044 1654	0·00047 000
29	0·03448 27586	0·00333 87120	0·00092 87651	0·00046 26319	0·00034 4771	0·00034 921
30	0·03333 33333	0·00305 46077	0·00080 67867	0·00038 22384	0·00027 12811	0·00026 1932
31	0·03225 80645	0·00280 25904	0·00070 39300	0·00031 76966	0·00021 50445	0·00019 8216
32	0·03125 00000	0·00257 81889	0·00061 67367	0·00026 52568	0·00017 16569	0·00015 1249
33	0·03030 30303	0·00237 76700	0·00054 24524	0·00022 30929	0·00013 79231	0·00011 6318
34	0·02941 17647	0·00219 78937	0·00047 88690	0·00018 83671	0·00011 15029	0·00009 0117
35	0·02857 14285	0·00203 62079	0·00042 42044	0·00015 97894	0·00009 06725	0·00007 03046
36	0·02777 77778	0·00189 03610	0·00037 70113	0·00013 61459	0·00007 41397	0·00005 51924
37	0·02702 70270	0·00175 84330	0·00033 61078	0·00011 64857	0·00006 09405	0·00004 36264
38	0·02631 57894	0·00163 87790	0·00030 05231	0·00010 00595	0·00005 03360	0·00003 46801
39	0·02564 10256	0·00152 99842	0·00026 94559	0·00008 62731	0·00004 17740	0·00002 77157
40	0·02500 00000	0·00143 08270	0·00024 22414	0·00007 46525	0·00003 48236	0·00002 22795
41	0·02439 02439	0·00134 02485	0·00021 83255	0·00006 48171	0·00002 91534	0·00001 79963
42	0·02380 95238	0·00125 73278	0·00019 72443	0·00005 64603	0·00002 45055	0·00001 46075
43	0·02325 58139	0·00118 12613	0·00017 86078	0·00004 93332	0·00002 06786	0·00001 19117
44	0·02272 72727	0·00111 13455	0·00016 20870	0·00004 32335	0·00001 75139	0·00000 97565
45	0·02222 22222	0·00104 69623	0·00014 74026	0·00003 79951	0·00001 48862	0·00000 80238
46	0·02173 91304	0·00098 75678	0·00013 43177	0·00003 34819	0·00001 26956	0·00000 66278
47	0·02127 65957	0·00093 26812	0·00012 26296	0·00002 95815	0·00001 08626	0·00000 54951
48	0·02083 33333	0·00088 18772	0·00011 21650	0·00002 62005	0·00000 93233	0·00000 45730
49	0·02040 81632	0·00083 47778	0·00010 27750	0·00002 32614	0·00000 80260	0·00000 38192
50	0·02000 00000	0·00079 10470	0·00009 43311	0·00002 06994	0·00000 69292	0·00000 32007

† For convenience, we have written $\bar{V}_{\beta,\beta}(x) = \bar{u}_\beta(x)$.

Table 1 (*cont.*)

n	$\bar{u}_{n-2}(3)$	$\bar{u}_{n-4}(5)$	$\bar{u}_{n-6}(7)$	$\bar{u}_{n-8}(9)$	$\bar{u}_{n-10}(11)$
1	—	—	—	—	—
2	1·00000 00000	—	—	—	—
3	0·50000 00000	—	—	—	—
4	0·30408 67240	1·00000 00000	—	—	—
5	0·20613 00860	0·50000 00000	—	—	—
6	0·14973 76529	0·28204 71084	1·00000 000	—	—
7	0·11412 73223	0·17307 06626	0·50000 000	—	—
8	0·09011 66373	0·11301 25446	0·27281 447	1·00000 00	—
9	0·07311 43694	0·07741 35906	0·15922 171	0·50000 00	—
10	0·06060 81823	0·05508 32201	0·09803 672	0·26772 05	1·00000 0
11	0·05112 51858	0·04042 61431	0·06305 560	0·15158 07	0·50000 0
12	0·04375 34970	0·03044 14257	0·04205 400	0·08990 27	0·26448 9
13	0·03790 30207	0·02342 53126	0·02891 971	0·05545 54	0·14673 3
14	0·03317 76395	0·01836 38033	0·02041 603	0·03537 66	0·08479 5
15	0·02930 31093	0·01462 89404	0·01474 400	0·02323 49	0·05076 6
16	0·02608 44854	0·01181 83990	0·01086 154	0·01565 44	0·03135 0
17	0·02337 99954	0·00966 67260	0·00814 300	0·01078 73	0·01989 5
18	0·02108 44609	0·00799 42195	0·00620 083	0·00758 37	0·01293 8
19	0·01911 84699	0·00667 64417	0·00478 826	0·00542 79	0·00859 9
20	0·01742 11230	0·00562 54722	0·00374 4257	0·00394 826	0·00582 81
21	0·01594 50617	0·00477 80637	0·00296 1394	0·00291 427	0·00402 11
22	0·01465 29936	0·00408 79746	0·00236 6593	0·00217 989	0·00281 97
23	0·01351 52178	0·00352 08978	0·00190 9238	0·00165 053	0·00200 67
24	0·01250 78348	0·00305 10386	0·00155 3697	0·00126 376	0·00144 76
25	0·01161 14315	0·00265 87653	0·00127 4511	0·00097 762	0·00105 75
26	0·01081 01025	0·00232 89690	0·00105 3421	0·00076 350	0·00078 15
27	0·01009 07126	0·00204 99013	0·00087 6364	0·00060 155	0·00058 37
28	0·00944 23349	0·00181 23400	0·00073 3843	0·00047 785	0·00044 04
29	0·00885 58175	0·00160 89817	0·00061 8152	0·00038 249	0·00033 54
30	0·00832 34467	0·00143 39955	0·00052 35919	0·00030 8361	0·00025 762
31	0·00783 86833	0·00128 26902	0·00044 58018	0·00025 0263	0·00019 951
32	0·00739 59526	0·00115 12650	0·00038 14199	0·00020 4389	0·00015 569
33	0·00699 04786	0·00103 66196	0·00032 78320	0·00016 7911	0·00012 237
34	0·00661 81492	0·00093 62093	0·00028 29897	0·00013 8712	0·00009 684
35	0·00627 54084	0·00084 79336	0·00024 52764	0·00011 5194	0·00007 712
36	0·00595 91679	0·00077 00483	0·00021 34076	0·00009 6139	0·00006 180
37	0·00566 67346	0·00070 10983	0·00018 63565	0·00008 0613	0·00004 980
38	0·00539 57516	0·00063 98631	0·00016 32967	0·00006 7896	0·00004 035
39	0·00514 41482	0·00058 53144	0·00014 35599	0·00005 7428	0·00003 286
40	0·00491 00998	0·00053 65820	0·00012 66023	0·00004 8770	0·00002 689
41	0·00469 19927	0·00049 29266	0·00011 79793	0·00004 1576	0·00002 211
42	0·00448 83958	0·00045 37172	0·00009 93256	0·00003 5572	0·00001 827
43	0·00429 80362	0·00041 84134	0·00008 83397	0·00003 0542	0·00001 515
44	0·00411 97785	0·00038 65508	0·00007 87716	0·00002 6311	0·00001 262
45	0·00395 26071	0·00035 77288	0·00007 04131	0·00002 2738	0·00001 055
46	0·00379 56117	0·00033 16006	0·00006 30901	0·00001 9710	0·00000 885
47	0·00364 79741	0·00030 78655	0·00005 66566	0·00001 7136	0·00000 746
48	0·00350 89572	0·00028 62613	0·00005 09895	0·00001 4939	0·00000 630
49	0·00337 78955	0·00026 65593	0·00004 59848	0·00001 3059	0·00000 534
50	0·00325 41871	0·00024 85591	0·00004 15543	0·00001 1446	0·00000 454

Table 2. *Moments of extreme members, $\mu'_{n|n}$, in samples of size n drawn from normal populations*

n \ r	1	2	3	4	5
1	0	1·00000 0000	0	3·00000 0000	0
2	0·56418 9584	1·00000 0000	1·41047 3959	3·00000 0000	6·06503 8023
3	0·84628 4377	1·27566 4448	2·11571 0938	4·19454 5940	9·09755 7036
4	1·02937 5374	1·55132 8896	2·70042 5704	5·38909 1881	11·88062 0268
5	1·16296 4474	1·80002 0436	3·22487 9364	6·52339 5486	14·53895 5578
6	1·26720 6361	2·02173 9069	3·70526 1794	7·59745 6758	17·09454 9080
7	1·35217 8376	2·22030 4136	4·14966 7934	8·61704 4920	19·55839 3798
8	1·42360 0306	2·39953 4975	4·56358 1597	9·58792 9198	21·93884 615
9	1·48501 3162	2·56261 7418	4·95118 1032	10·51515 5116	24·24294 426
10	1·53875 2731	2·71210 3790	5·31580 4079	11·40304 4507	26·47678 168
11	1·58643 6352	2·85002 7742	5·66018 0737	12·25530 1015	28·64569 426
12	1·62922 7640	2·97801 9090	5·98657 7532	13·07511 5628	30·75438 816
13	1·66799 0177	3·09739 6615	6·29689 7766	13·86525 5250	32·80703 849
14	1·70338 1554	3·20923 8822	6·59275 5132	14·62813 4342	34·80736 837
15	1·73591 3445	3·31443 7059	6·87552 9415	15·36587 2528	36·75871 365
16	1·76599 1393	3·41373 5410	7·14640 9450	16·08034 1083	38·66407 622
17	1·79394 1981	3·50776 0835	7·40642 6717	16·77320 0671	40·52616 843
18	1·82003 1879	3·59704 6171	7·65648 1911	17·44593 2213	42·34745 031
19	1·84448 1512	3·68204 7852	7·89736 6170	18·09986 2296	44·13016 084
20	1·86747 5060	3·76315 9715	8·12977 8192	18·73618 4245	45·87634 459
21	1·88916 7917	3·84072 3854	8·35433 8148	19·35597 5699	47·58787 434
22	1·90969 2325	3·91503 9251	8·57159 9074	19·96021 3381	49·26647 044
23	1·92916 1713	3·98636 8684	8·78205 6290	20·54978 5557	50·91371 752
24	1·94767 4075	4·05494 4325	8·98615 5205	21·12550 263	52·53107 895
25	1·96531 4610	4·12097 2294	9·18429 7863	21·68810 616	54·11990 930
26	1·98215 7840	4·18463 6408	9·37684 8436	22·23827 665	55·68146 535
27	1·99826 9303	4·24610 1279	9·56413 7882	22·77664 017	57·21691 556
28	2·01370 6925	4·30551 4889	9·74646 7901	23·30377 420	58·72734 852
29	2·02852 2147	4·36301 0759	9·92411 4318	23·82021 261	60·21378 036
30	2·04276 0845	4·41870 9768	10·09732 9992	24·32645 008	61·67716 127
31	2·05646 4099	4·47272 1701	10·26634 7323	24·82294 588	63·11838 132
32	2·06966 8829	4·52514 6562	10·43138 0426	25·31012 735	64·53827 561
33	2·08240 8337	4·57607 5706	10·59262 7020	25·78839 276	65·93762 894
34	2·09471 2757	4·62559 2804	10·75027 0082	26·25811 404	67·31717 988
35	2·10660 9441	4·67377 4683	10·90447 9295	26·71963 905	68·67762 449
36	2·11812 3288	4·72069 2054	11·05541 2315	27·17329 371	70·01961 969
37	2·12927 7027	4·76641 0142	11·20321 5903	27·61938 383	71·34378 620
38	2·14009 1456	4·81098 9241	11·34802 6914	28·05819 679	72·65071 130
39	2·15058 5659	4·85448 5197	11·48997 3178	28·49000 300	73·94095 130
40	2·16077 7179	4·89694 9839	11·62917 4294	28·91505 728	75·21503 372
41	2·17068 2186	4·93843 1351	11·76574 2323	29·33360 007	76·47345 938
42	2·18031 5609	4·97897 4611	11·89978 2420	29·74585 847	77·71670 421
43	2·18969 1262	5·01862 1481	12·03139 3399	30·15204 729	78·94522 094
44	2·19882 1950	5·05741 1077	12·16066 8233	30·55236 992	80·15944 068
45	2·20771 9565	5·09538 0002	12·28769 4522	30·94701 916	81·35977 429
46	2·21639 5169	5·13256 2557	12·41255 4895	31·33617 792	82·54661 372
47	2·22485 9069	5·16899 0930	12·53532 7387	31·72001 998	83·72033 315
48	2·23312 0882	5·20469 5366	12·65608 5780	32·09871 054	84·88129 016
49	2·24118 9599	5·23970 4322	12·77489 9908	32·47240 679	86·02982 667
50	2·24907 3631	5·27404 4605	12·89183 5941	32·84125 849	87·16626 994

Table 2 (*cont.*)

n \ r	6	7	8	9	10
1	15·00000 000	0	105·00000 0	0	945·00000
2	15·00000 000	37·44808 36	105·00000 0	303·28716 3	945·00000
3	21·79972 305	56·17212 54	155·23218 8	454·93074 5	1407·98355
4	28·59944 610	74·16003 19	205·46437 7	603·73513 9	1870·96710
5	35·22660 772	91·77987 07	255·06652 6	751·11993 7	2331·11618
6	41·68120 792	109·07401 19	304·03863 8	897·19147 1	2788·43082
7	47·97432 643	126·06364 04	352·40762 6	1042·00290 7	3242·99105
8	54·11704 295	142·76617 39	400·20040 7	1185·60041 6	3694·87692
9	60·11939 069	159·19714 68	447·44198 8	1328·02667 6	4144·16423
10	65·99035 636	175·37063 38	494·15547 2	1469·32149 3	4590·92456
11	71·73800 254	191·29946 19	540·36223 3	1609·52211 3	5035·22556
12	77·36959 015	206·99536 78	586·08209 7	1748·66346 5	5477·13127
13	82·89168 443	222·46912 62	631·33349 6	1886·77837 6	5916·70244
14	88·31024 450	237·73065 92	676·13361 2	2023·89775 2	6353·99673
15	93·63069 901	252·78913 03	720·49849 5	2160·05074 7	6789·06900
16	98·85801 015	267·65302 48	764·44316 9	2295·26490 1	7221·97147
17	103·99672 809	282·33021 97	807·98172 9	2429·56628 0	7652·75394
18	109·05103 766	296·82804 46	851·12741 8	2562·97958 8	8081·46396
19	114·02479 836	311·15333 50	893·89270 6	2695·52826 2	8508·14696
20	118·92157 897	325·31247 89	936·28934 9	2827·23458 7	8932·84640
21	123·74468 755	339·31145 76	978·32845 4	2958·11976 4	9355·60392
22	128·49719 757	353·15588 21	1020·02052 5	3088·20399 1	9776·45944
23	133·18197 078	366·85102 56	1061·37551 5	3217·50653 5	10195·45124
24	137·80167 727	380·40185 19	1102·40286 1	3346·04579 3	10612·61612
25	142·35881 312	393·81304 07	1143·11152 9	3473·83935 1	11027·98943
26	146·85571 609	407·08901 08	1183·51004 4	3600·90403 7	11441·60519
27	151·29457 942	420·23394 06	1223·60652 3	3727·25596 8	11853·49616
28	155·67746 420	433·25178 61	1263·40870 2	3852·91059 3	12263·69389
29	160·00631 037	446·14629 83	1302·92396 2	3977·88273 7	12672·22882
30	164·28294 657	458·92103 76	1342·15935 3	4102·18663 7	13079·13032
31	168·50909 90	471·57938 78	1381·12161 5	4225·83597 5	13484·42672
32	172·68639 92	484·12456 81	1419·81719 7	4348·84391 1	13888·14543
33	176·81639 16	496·55964 50	1458·25227 8	4471·22311 0	14290·31294
34	180·90053 95	508·88754 19	1496·43277 8	4592·98577 6	14690·95486
35	184·94023 16	521·11104 91	1534·36438 0	4714·14366 8	15090·09599
36	188·93678 64	533·23283 20	1572·05253 8	4834·70813 0	15487·76036
37	192·89145 81	545·25543 90	1609·50249 4	4954·69011 0	15883·97122
38	196·80544 02	557·18130 91	1646 71928 9	5074·10018 2	16278·75114
39	200·67987 00	569·01277 81	1683·70777 2	5192·94856 3	16672·12201
40	204·51583 21	580·75208 53	1720·47261 3	5311·24513 1	17064·10506
41	208·31436 17	592·40137 87	1757·01831 5	5428·99944 4	17454·72091
42	212·07644 81	603·96272 02	1793·34921 5	5546·22075 1	17843·98959
43	215·80303 72	615·43809 10	1829·46950 0	5662·91801 2	18231·93056
44	219·49503 42	626·82939 59	1865·38321 0	5779·09990 6	18618·56273
45	223·15330 61	638·13846 69	1901·09424 9	5894·77474 8	19003·90451
46	226·77868 38	649·36706 80	1936·60639 0	6009·95099 8	19387·97380
47	230·37196 46	660·51689 81	1971·92327 9	6124·63627 3	19770·78803
48	233·93391 36	671·58959 50	2007·04844 5	6238·83836 1	20152·36417
49	237·46526 56	682·58673 78	2041·98530 5	6352·56472 5	20532·71874
50	240·96672 72	693·50985 06	2076·73716 4	6465·82261 9	20911·86784

Table 3. *The moments about the mean, the measures of skewness and of kurtosis and the standard deviation of extreme members in samples of size n drawn from normal populations*

n	μ_2	μ_3	μ_4	β_1	β_2-3	$\sqrt{\mu_2}$
1	1·00000 000	0	3·00000 00	0	0	1·00000 00
2	0·68169 011	0·07707 945	1·42279 69	0·01875 50	0·06174 43	0·82564 53
3	0·55946 720	0·08919 934	0·97955 22	0·04543 59	0·12952 43	0·74797 54
4	0·49171 524	0·09120 683	0·76459 74	0·06997 03	0·16231 77	0·70122 41
5	0·44753 407	0·09058 710	0·64107 54	0·09154 92	0·20078 82	0·66897 99
6	0·41592 711	0·08917 023	0·55943 20	0·11050 66	0·23379 78	0·64492 41
7	0·39191 778	0·08753 521	0·50113 26	0·12728 60	0·26259 20	0·62603 34
8	0·37289 714	0·08588 990	0·45721 20	0·14227 12	0·28805 84	0·61065 30
9	0·35735 333	0·08431 165	0·42279 90	0·15576 92	0·31084 08	0·59779 04
10	0·34434 382	0·08282 697	0·39501 52	0·16802 22	0·33141 89	0·58680 82
11	0·33324 744	0·08144 141	0·37204 79	0·17922 15	0·35015 75	0·57727 59
12	0·32363 639	0·08015 182	0·35269 69	0·18951 99	0·36733 95	0·56889 05
13	0·31520 538	0·07895 167	0·33613 48	0·19904 06	0·38318 91	0·56143 15
14	0·30773 010	0·07783 336	0·32177 24	0·20788 42	0·39788 63	0·55473 43
15	0·30104 157	0·07678 934	0·30917 77	0·21613 37	0·41157 72	0·54867 26
16	0·29500 981	0·07581 253	0·29802 68	0·22385 80	0·42438 30	0·54314 81
17	0·28953 300	0·07489 647	0·28807 16	0·23111 52	0·43640 41	0·53808 27
18	0·28453 013	0·07403 540	0·27911 89	0·23795 45	0·44772 53	0·53341 37
19	0·27993 580	0·07322 415	0·27101 57	0·24441 81	0·45841 87	0·52908 96
20	0·27569 662	0·07245 817	0·26363 94	0·25054 19	0·46854 62	0·52506 82
21	0·27176 844	0·07173 345	0·25689 02	0·25635 77	0·47816 00	0·52131 41
22	0·26811 447	0·07104 635	0·25068 64	0·26189 23	0·48730 80	0·51779 77
23	0·26470 377	0·07039 370	0·24496 01	0·26716 97	0·49603 01	0·51449 37
24	0·26151 002	0·06977 268	0·23965 45	0·27221 11	0·50436 15	0·51138 05
25	0·25851 078	0·06918 079	0·23472 16	0·27703 53	0·51233 28	0·50843 96
26	0·25568 671	0·06861 576	0·23012 06	0·28165 88	0·51997 21	0·50565 47
27	0·25302 107	0·06807 556	0·22581 68	0·28609 63	0·52730 46	0·50301 20
28	0·25049 931	0·06755 837	0·22178 02	0·29036 13	0·53435 19	0·50049 91
29	0·24810 866	0·06706 255	0·21798 48	0·29446 55	0·54113 42	0·49810 51
30	0·24583 790	0·06658 663	0·21440 79	0·29841 97	0·54766 97	0·49582 04
31	0·24367 711	0·06612 926	0·21102 97	0·30223 38	0·55397 34	0·49363 66
32	0·24161 750	0·06568 923	0·20783 28	0·30591 67	0·56006 08	0·49154 60
33	0·23965 122	0·06526 541	0·20480 20	0·30947 61	0·56594 66	0·48954 19
34	0·23777 127	0·06485 680	0·20192 34	0·31291 97	0·57164 13	0·48761 80
35	0·23597 135	0·06446 249	0·19918 49	0·31625 42	0·57715 59	0·48576 88
36	0·23424 579	0·06408 161	0·19657 59	0·31948 58	0·58250 37	0·48398 95
37	0·23258 948	0·06371 339	0·19408 62	0·32262 00	0·58768 62	0·48227 53
38	0·23099 780	0·06335 710	0·19170 75	0·32566 20	0·59272 11	0·48062 23
39	0·22946 652	0·06301 212	0·18943 18	0·32861 71	0·59761 04	0·47902 66
40	0·22799 182	0·06267 778	0·18725 20	0·33148 92	0·60236 66	0·47748 49
41	0·22657 020	0·06235 357	0·18516 15	0·33428 30	0·60699 15	0·47599 39
42	0·22519 846	0·06203 894	0·18315 46	0·33700 20	0·61149 41	0·47455 08
43	0·22387 366	0·06173 341	0·18122 58	0·33964 99	0·61587 96	0·47315 29
44	0·22259 311	0·06143 652	0·17937 03	0·34222 99	0·62015 48	0·47179 77
45	0·22135 432	0·06114 786	0·17758 36	0·34474 54	0·62432 34	0·47048 31
46	0·22015 501	0·06086 702	0·17586 17	0·34719 89	0·62839 06	0·46920 68
47	0·21899 305	0·06059 368	0·17420 05	0·34959 37	0·63235 93	0·46796 69
48	0·21786 649	0·06032 745	0·17259 69	0·35193 19	0·63623 67	0·46676 17
49	0·21677 350	0·06006 804	0·17104 75	0·35421 62	0·64002 38	0·46558 94
50	0·21571 241	0·05981 512	0·16954 94	0·35644 83	0·64372 80	0·46444 85

REFERENCES

COLE, R. H. (1951). Relations between moments of order statistics. *Ann. Math. Statist.* **22**, 308–10.
COXETER, H. S. M. (1935). The functions of Schläfli and Lobachewski. *Quart. J. Math.* **6**, 13–29.
COXETER, H. S. M. (1948). *Regular Polytopes*. London: Methuen and Co. Ltd.
GODWIN, H. (1949*a*). Some low moments of order statistics. *Ann. Math. Statist.* **20**, 279–85.
GODWIN, H. (1949*b*). On the estimation of dispersion by linear systematic statistics. *Biometrika*, **36**, 92–100.
HASTINGS, C., Jr., MOSTELLER, F., TUKEY, J. W. & WINSOR, C. P. (1947). Low moments for small samples: a comparative study of order statistics. *Ann. Math. Statist.* **18**, 413–26.
HOPPE, E. R. E. (1882). Berechnung einiger viehrdehnigen Winkel. *Arch. Math. Phys.* **67**, 269–90.
HOJO, T. (1931). Distribution of the median, quartiles and interquartile distance in samples from a normal population. *Biometrika*, **23**, 315–60.
JONES, H. L. (1948). Exact lower moments of order statistics in small samples from a normal distribution. *Ann. Math. Statist.* **19**, 270–3.
RICHMOND, H. W. (1903). The volume of a tetrahedron in elliptic space. *Quart. J. Pure Appl. Math.* **34**, 175–7.
SCHLÄFLI, L. (1950). *Gesammelte Mathematische Abhandlungen*, Vol. 1. (Lehre von der sphärischen Kontinuen, 227–98). Verlag Birkhäuser Basel.
SOMMERVILLE, D. M. Y. (1929). *An Introduction to the Geometry of N Dimensions*. London: Methuen and Co. Ltd.
TIPPETT, L. H. C. (1925). On the extreme individuals and the range of samples taken from a normal population. *Biometrika*, **17**, 364.

† Published posthumously, written in 1850–2. Part translated by Cayley (1858–60) *Quart. J. Math.* **2**, 269–301; **3**, 54–68, 97–108.

STATISTICAL TREATMENT OF CENSORED DATA

PART I. FUNDAMENTAL FORMULAE

By F. N. DAVID and N. L. JOHNSON
University College, London

1. In recently published papers a distinction has been made between samples which are said to be 'censored' and those which are called 'truncated'. In the former case the total number of observations is known but full information is not available in regard to some of them. In the latter case not only is information lacking in regard to part of the sample but the total number of observations is not known.

In this paper *censored* data will be considered, and for ease of presentation we shall consider only samples censored on the right. This means that the only information about the missing observations is their total number and the fact that each is greater than some known value.

2. The methods which we intend to develop are based on the use of ordered variates. This is a natural way of approach, since in a censored distribution the rank of each observation available is known.

In the present paper we are concerned primarily with a number of basic formulae which will be used in later work. Some applications are described, but they are of simple type and restricted to the subclass of censored distributions wherein the method of censoring is such that the number of omitted values is fixed in advance and not determined by the number of observations which happen to be greater than some predetermined value. This means that in repeated sampling we will always have the same ordered variates available. Such situations arise, for example, in life-testing, if survival times of the first k individuals to die out of N are observed, as opposed to the system wherein all survival times up to a fixed *duration* are observed. (Gupta (1952) has classified these as Type I and Type II censoring respectively.)

Cases of the latter type, and cases involving censoring of a more complicated type, will be considered in later papers.

3. It will be assumed that there is a population of measurements which can be described by a continuous random variable with probability density function $f(t)$. We shall write

$$F(X) = \int_{-\infty}^{X} f(t)\,dt. \tag{1}$$

Assume that a (complete) random sample of n observations has been drawn from this population. Let these observations, arranged in ascending order of magnitude, correspond to the random variables $x_1, x_2, \ldots, x_n$. Let $X_1, X_2, \ldots, X_n$ be defined by the equations

$$F(X_r) = r/(n+1) \quad (r = 1, \ldots, n). \tag{2}$$

We will then formally expand x_r about X_r in an inverse Taylor series, obtaining

$$x_r = X_r + X_r' h(x_r) + \tfrac{1}{2} X_r''[h(x_r)]^2 + \ldots, \tag{3}$$

where $h(x_r) = F(x_r) - F(X_r) = F(x_r) - r/(n+1)$

and $$X'_r = \frac{dX_r}{dF} = \frac{dX}{dF}\bigg|_{X=X_r}, \quad X''_r = \frac{d^2X_r}{dF^2} = \frac{d^2X}{dF^2}\bigg|_{X=X_r},$$

and so on. This is a modification of a method of approach devised by K. Pearson (1931).

It will not always be legitimate to expand in this way. In particular, if n is large and r is near either n or 1, convergence may be slow or non-existent. However, for the cases considered in this paper this expansion gives useful results. In the case of fourth-order cumulants we have given only the leading term in each expansion, and special calculations may sometimes be necessary to obtain an appreciation of their accuracy.

4. The inverse Taylor expansion enables us to find approximations to the moments of the x_r's from the moments of the $h(x_r)$'s, and the latter moments are known. For example, the joint probability density function of $F(x_r)$, $F(x_s)$, $F(x_t)$ and $F(x_v)$ $(r<s<t<v)$ is

$$p(F_r, F_s, F_t, F_v) = \frac{n!}{(r-1)!\,(s-r-1)!\,(t-s-1)!\,(v-t-1)!\,(n-v)!} \times F_r^{r-1}(F_s-F_r)^{s-r-1}(F_t-F_s)^{t-s-1}(F_v-F_t)^{v-t-1}(1-F_v)^{n-v},$$

where $F_r \equiv F(x_r)$ and so on.

It follows that

$$\mathscr{E}[F_r^l F_s^m F_t^q F_v^h] = \frac{n!}{(n+l+m+g+h)!}\,\frac{(r+l-1)!}{(r-1)!}\,\frac{(s+l+m-1)!}{(s+l-1)!}\,\frac{(t+l+m+g-1)!}{(t+l+m-1)!}\,\frac{(v+l+m+g+h-1)!}{(v+l+m+g-1)!}.$$

In general

$$\mathscr{E}\left[\prod_{i=1}^{N} F_{r_i}^{l_i}\right] = \frac{n!}{\left(n+\sum_{i=1}^{N} l_i\right)!}\prod_{j=1}^{N}\frac{\left(r_j+\sum_{i=1}^{j} l_i-1\right)!}{\left(r_j+\sum_{i=1}^{j-1} l_i-1\right)!}. \tag{4}$$

Central moments and cumulants derived from (4) are shown in Table 1. In this table it was found convenient to use the symbols $p_r = r/(n+1)$ and $q_r = 1-p_r$.

Combining these results with (3) we obtain formal expansions for the moments of x_r. Care is necessary to include all terms of appropriate order in the expansion. The variable form of the denominator in (4) leaves some choice in the form of expansion to use. The expansion may be in powers of n^{-1} or $(n+1)^{-1}$, or $(n+2)^{-2}$ and so on. In Table 2 we have chosen expansion in powers of $(n+2)^{-1}$ for the following reasons:

(i) It has the simplest form for the first two moments.

(ii) It gives the closest approximation to the exact results of Hojo (1931) for the expected value of quartiles from Normal populations.

(iii) It gives exact results for the first two moments of any ordered variate when the population is rectangular.

These are not, of course, at all conclusive arguments, and in many cases some other function of n may give a more quickly convergent expansion.

5. We will now digress to give a few interesting results obtainable from Table 2. This table gives approximate expressions for the moments and product-moments of ordered variates in samples from any population for which there is a probability-density function which is differentiable as many times as desired. Some special cases will now be considered.

(i) *Median.* Assuming n is odd and writing $m = \frac{1}{2}(n+1)$, we have, to the appropriate order in $(n+2)^{-1}$,

$$\mathscr{E}(x_m) = X_m + \frac{1}{8(n+2)} X''_m + \frac{1}{128(n+2)^2} X^{iv}_m,$$

$$\kappa(x^2_m) = \frac{1}{4(n+2)} X'^2_m + \frac{1}{32(n+2)^2}(2X'_m X'''_m + X''^2_m),$$

$$\kappa(x^3_m) = \frac{3}{16(n+2)^2} X'^2_m X''_m,$$

$$\kappa(x^4_m) = \frac{1}{16(n+2)^3}(3X'^2_m X''^2_m + X'^3_m X'''_m - 6X'^4_m),$$

and

$$\gamma_1(x_m) = \frac{\kappa(x^3_m)}{[\kappa(x^2_m)]^{\frac{3}{2}}} = \frac{3}{2\sqrt{(n+2)}} \frac{X''_m}{X'_m},$$

$$\gamma_2(x_m) = \frac{\kappa(x^4_m)}{[\kappa(x^2_m)]^2} = \frac{1}{(n+2)}\left(\frac{X'''_m}{X'_m} + \frac{3X''^2_m}{X'^2_m} - 6\right).$$

Using the inverse Edgeworth (or Cornish-Fisher, 1937) expansion, formal expansions for percentage points of the distribution of the median may be obtained. Thus, if $x'_{m,\alpha}$ be the standardized upper $100\alpha\,\%$ point of the distribution of the median, i.e.

$$P\{x_m - \mathscr{E}(x_m) > x'_{m,\alpha}[\kappa(x^2_m)]^{\frac{1}{2}}\} = \alpha,$$

then, to order $(n+2)^{-1}$

$$x'_{m,\alpha} = \lambda_\alpha + \frac{1}{4\sqrt{(n+2)}} \frac{X''_m}{X'_m}(\lambda^2_\alpha - 1) + \frac{1}{48(n+2)}\left[2\left(\frac{X'''_m}{X'_m} - 6\right)(\lambda^3_\alpha - 3\lambda_\alpha) - \frac{3X''^2_m}{X'^2_m}\lambda_\alpha\right],$$

where λ_α is the upper $100\alpha\,\%$ point of the unit Normal distribution.

Results for four special forms of distribution are summarized below.

Name	Probability density function	X'_m	X''_m	X'''_m	$\kappa(x^2_m)$	$x'_{m,\alpha} - \lambda_\alpha$
Normal	$\frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$	$(2\pi)^{\frac{1}{2}}$	0	$(2\pi)^{\frac{3}{2}}$	$\frac{\pi}{2(n+2)} + \frac{\pi^2}{4(n+2)^2}$	$\frac{\pi-3}{12(n+2)}(\lambda^3_\alpha - 3\lambda_\alpha)$
Rectangular	$1\ (0<x<1)$	1	0	0	$\frac{1}{4}(n+2)^{-1}$	$-\frac{1}{4}(n+2)^{-1}(\lambda^3_\alpha - 3\lambda_\alpha)$
Cauchy	$\pi^{-1}(1+x^2)^{-1}$	π	0	$2\pi^3$	$\frac{1}{2}\pi^2(n+2)^{-1} + \frac{1}{3}\pi^4(n+2)^{-2}$	$-\frac{1}{12}(\pi^2-3)(n+2)^{-1}(\lambda^3_\alpha - 3\lambda_\alpha)$
Exponential	$e^{-x}\ (x>0)$	2	4	16	$(n+2)^{-1} + \frac{5}{4}(n+2)^{-2}$	$\frac{1}{2}(n+2)^{-1}(\lambda^2_\alpha - 1) + \frac{1}{12}(n+2)^{-1}(\lambda^3_\alpha - 6\lambda_\alpha)$

(As noted above, the moments for the rectangular distribution are exact.)

(ii) *Correlation between ordered variates.* To order $(n+2)^{-1}$ the formal expression for the correlation between x_r and x_s $(r<s)$ is

$$\rho(r,s) = \sqrt{\left(\frac{p_r q_s}{p_s q_r}\right)}\left[1 - \frac{1}{4(n+2)}\left\{\frac{X''^2_r}{X'^2_r} p_r q_r - \frac{2X''_r X''_s}{X'_r X'_s} p_r q_s + \frac{X''^2_s}{X'^2_s} p_s q_s\right\}\right].$$

The following table compares values of $\rho(r, s)$ for a Normal population calculated from the above approximation with exact values given by Godwin (1949).

Values of $\rho(r, s)$

n	r	s	Approx.	Exact
3	1	3	0·2958	0·2947
4	1	4	0·2123	0·2129
	2	3	0·6571	0·6546
9	1	9	0·0842	0·0869
	2	8	0·2295	0·2291
	3	7	0·4153	0·4144
	4	6	0·6615	0·6606

(iii) *Quartiles and interquartile distance.* The upper quartile will be defined as x_u, where $u = \frac{3}{4}(n+1)$ and the lower quartile as x_l, where $l = \frac{1}{4}(n+1)$, so that $p_u = \frac{3}{4}$, $p_l = \frac{1}{4}$. For both u and l to be integers it is clear that $(n+1)$ must be a multiple of four. For a unit Normal population we find, to order $(n+2)^{-3}$,

$$\mathscr{E}(x_u) = 0{\cdot}67449 + \frac{0{\cdot}62617}{n+2} + \frac{0{\cdot}96810}{(n+2)^2} + \frac{1{\cdot}00287}{(n+2)^3}$$

and

$$\mathrm{var}\,(x_u) = \frac{1{\cdot}85677}{n+2} + \frac{3{\cdot}42763}{(n+2)^2} + \frac{5{\cdot}35871}{(n+2)^3}.$$

Comparisons with Hojo's exact values are shown below.

	Expectation		Variance	
n	Approx.	Exact	Approx.	Exact
7	0·75739	0·75737	0·25598	0·25673
11	0·72885	0·72885	0·16555	0·16571

For the exponential population we have, to order $(n+2)^{-3}$,

$$\mathscr{E}(x_u) = \log_e 4 + \frac{3}{2(n+2)} + \frac{11}{4(n+2)^2} + \frac{4}{(n+2)^3},$$

$$\mathscr{E}(x_l) = \log_e \frac{4}{3} + \frac{1}{6(n+2)} + \frac{25}{108(n+2)^2} + \frac{8}{27(n+2)^3}.$$

When $n = 11$ we find that these formulae give $\mathscr{E}(x_u) = 1{\cdot}51977$, $\mathscr{E}(x_l) = 0{\cdot}30201$, as compared with exact values 1·51988 and 0·30202 respectively.

The expected value of the interquartile distance $(I = x_u - x_l)$ is, of course, $\mathscr{E}(x_u) - \mathscr{E}(x_l)$. The variance and higher moments of I are obtained by straightforward application of results in Table 2.

6. We will now consider the problem of testing certain simple hypotheses, restricting ourselves to data censored on the right. We will suppose that the lowest k of a random sample of n observations are recorded. Since we are assuming that the lowest k observations are always available we can refer them to the distributions for which we have obtained formal expansions for moments and product-moments. In the case of a simple hypothesis we can make use of the probability-integral transformation to reduce our problem to that of dealing with a rectangular population with consequent simplification of many of our formulae.

Let the available observations in ascending order be $t_1, t_2, \ldots, t_k$, and denote by $f(t)$ the common probability-density function specified by the simple hypothesis H. Then we can apply the probability integral transformation

$$x_r = \int_{-\infty}^{t_r} f(t)\,dt.$$

If H is valid then the variates $x_1, \ldots, x_k$ will be distributed as ordered variates from the rectangular population $p(x) = 1\ (0 < x < 1)$.

If the alternative hypotheses simply specify a change in a pure location parameter, a natural criterion to use (if available) is the median, t_m, of the original observations. In many cases this might well be preferable to using x_m, but x_m does possess the advantage of generality.

When n is odd the upper $100\alpha\,\%$ point $x_{m,\alpha}$ of the distribution of x_m is given, in terms of the Incomplete beta-function ratio, by the equation

$$I_{x_{m,\alpha}}(\tfrac{1}{2}(n+1), \tfrac{1}{2}(n+1)) = 1-\alpha.$$

A few values of $x_{m,\alpha}$ are given below. Further values can be obtained as required from the tables of C. M. Thompson (1941)

n	$\alpha = 0{\cdot}05$	$\alpha = 0{\cdot}01$
5	0·811 (0·812)	0·894 (0·921)
15	0·700 (0·700)	0·767 (0·777)
25	0·659 (0·658)	0·719 (0·721)
Large	$\frac{1}{2}(1+1{\cdot}645(n+2)^{-\frac{1}{2}})$	$\frac{1}{2}(1+2{\cdot}326(n+2)^{-\frac{1}{2}})$

The values in brackets were calculated from the formula for $x'_{m,\alpha}$ given in § 5.

When n is even, the statistic $\tilde{x}_m = \frac{1}{2}(x_{\frac{1}{2}(n-1)} + x_{\frac{1}{2}(n+1)})$ may be used. An *approximate* equation for the upper $100\alpha\,\%$ point $\tilde{x}_{m,\alpha}$ is

$$I_{\tilde{x}_{m,\alpha}}(\tfrac{1}{2}n+1+n^{-1}, \tfrac{1}{2}n+1+n^{-1}) = 1-\alpha.$$

7. It will be recognized that the foregoing analysis is not comprehensive. It omits consideration of the situation arising if the median is not among the observed values, and does not take into account explicitly any observed values except the median.

In the notation of § 3 we have $X_r = r/(n+1) = \mathscr{E}(x_r)$ if the simple hypothesis H is valid. A change in the pure location parameter will tend to make the quantities $\mathscr{E}(x_r - X_r)$ all positive or all negative. We will therefore consider

$$\bar{x}_k = k^{-1} \sum_{i=1}^{k} (x_i - X_i) = k^{-1} \sum_{i=1}^{k} x_i - \tfrac{1}{2}(n+1)^{-1}(k+1)$$

as a possible criterion. We note that $\bar{x}_k$ has a finite range of distribution, from $-\frac{k+1}{2(n+1)}$ to $1-\frac{k+1}{2(n+1)}$. When H is valid the moments of $\bar{x}_k$ are given by the following equations:

$$\mathscr{E}(\bar{x}_k) = 0,$$

$$\kappa(\bar{x}_k^2) = k^{-2}(n+2)^{-1}\left[\sum_{i=1}^{k} p_i q_i + 2\sum_{i<j}^{k}\sum p_i q_j\right]$$

$$= k^{-2}(n+1)^{-2}(n+2)^{-1}\left[\sum_{i=1}^{k} i(n-i+1) + 2\sum_{i<j}^{k}\sum i(n-j+1)\right]$$

$$= \frac{k+1}{12k(n+1)^2(n+2)}[2(n+1)(2k+1) - 3k(k+1)],$$

$$\kappa(\bar{x}_k^3) = \frac{(k+1)^2(n-k+1)(n-k)}{2k(n+1)^3(n+2)(n+3)},$$

$$\kappa(\bar{x}_k^4) = \frac{k+1}{120k^3(n+1)^4(n+2)(n+3)(n+4)}[24(n+1)^3(2k+1)(3k^2+3k-1)$$

$$-40(n+1)^2k(k+1)(13k^2+13k+1) + 300(n+1)k^2(k+1)^2(2k+1)$$

$$-225k^3(k+1)^3 - 5(n+2)^{-1}k(k+1)\{2(n+1)(2k+1) - 3k(k+1)\}^2].$$

The evaluation of the above results involved the calculation of certain product sums of the first k natural numbers. Some of the results used were new to us in that we were not acquainted with any publication containing them. A summary of the results is appended in Table 3, and we shall use them in further work.

8. The distribution of $\bar{x}_k$ is remarkably skew when k is small compared with n. Values of β_1 and β_2 for the case $n = 10$ are shown in the following table:

k =	1	2	3	4	5	6	7	8	9	10
β_1 =	2·30	1·22	0·70	0·41	0·24	0·13	0·06	0·02	0·00	0·00
β_2 =	5·78	4·47	3·75	3·34	3·09	2·95	2·87	2·84	2·85	2·88

In view of the finite range of variation of $\bar{x}_k$ it was thought reasonable to consider using a Pearson Type I curve as an approximation to its distribution. Fitting a Type I curve to give the correct range of variation, mean and standard deviation, gave rather unsatisfactory results in that the third and fourth central moments of the fitted curve were much less than those of $\bar{x}_k$. Type I curves having correct values for the first four moments were therefore used, values of the standardized deviation of the upper and lower 5% points being obtained from the tables of Pearson & Merrington (1951). Values were also obtained using the Cornish-Fisher formula. The figures were in reasonable agreement with each other, though they cannot be relied upon to be sufficiently accurate for purposes of tabulation. They do, however, provide a standard of comparison for the approximation developed in the next section.

9. Consider the variable

$$y_k = \bar{x}_k + \tfrac{1}{2}(n+1)^{-1}(k+1) = k^{-1}\sum_{i=1}^{k} x_i.$$

The range of variation of y_k is from 0 to 1:

$$\mathscr{E}(y_k) = \frac{k+1}{2(n+1)} = P_k, \quad \text{say.}$$

$$\text{var}(y_k) = \tfrac{1}{3}(n+2)^{-1}P_k(2+k^{-1}-3P_k)$$

$$= \tfrac{1}{3}n^{-1}P_k(2-3P_k)F_k^2,$$

where F_k^n is a correcting factor which approaches unity as n and k increase (see table at end of this section).

Since $$\mathscr{E}(y_k) = P_k \quad \text{and} \quad \text{var}\,(y_k) \doteq \tfrac{1}{2}n^{-1}P_k(2-3P_k),$$

an appropriate variance equalizing transformation is

$$Y_k = \sin^{-1}\sqrt{\frac{3y_k}{2}} = \tfrac{1}{2}\cos^{-1}(1-3y_k) \quad (0<Y_k<\tfrac{1}{2}\pi).$$

Approximately $$\mathscr{E}(Y_k) = \tfrac{1}{2}\cos^{-1}(1-3P_k)$$

and $$\text{var}\,(Y_k) = \tfrac{1}{4}n^{-1}.$$

We note that if $y_k > \frac{2}{3}$ no corresponding value of Y_k can be obtained. It is to be expected, therefore, that the transformation will be useful only if $P\{y_k > \frac{2}{3}\}$ is small. Since

$$\mathscr{E}(y_k) = P_k < \tfrac{1}{2},$$

this will be the case for sufficiently large n.

Assuming Y_k to be normally distributed with mean and variance $\frac{1}{2}\cos^{-1}(1-3P_k)$ and $\frac{1}{4}n^{-1}$ respectively, we obtain the following upper and lower 5% points of $\bar{x}_k$. Values of F_k and approximate upper and lower 5% points obtained from the Cornish-Fisher expansion and from the fitted Type I are also shown in this table.

The use of the transformation as a basis for a quick significance test requires some further study in points of detail, but it appears likely to give reasonably useful practical results if n is bigger than 20 and $|\,kn^{-1}-\frac{1}{2}\,| < 0{\cdot}3$.

Percentage points of $\bar{x}_k$

n	k	Type I		Expansion		Transformation		F_k
		Lower 5%	Upper 5%	Lower 5%	Upper 5%	Lower 5%	Upper 5%	
10	2	−0·113†	0·171	−0·118	0·174	−0·108	0·160	1·047
	4	−0·142	0·182	−0·146	0·182	−0·143	0·171	0·996
	6	−0·160	0·179	−0·160	0·181	−0·163	0·168	0·983
	8	−0·161	0·169	−0·162	0·171	−0·171	0·151	0·984
20	4	−0·079	0·109	−0·079	0·109	−0·077	0·106	1·024
	8	−0·103	0·124	−0·103	0·123	−0·104	0·120	0·997
	12	−0·115	0·127	−0·115	0·127	−0·118	0·121	0·990
	16	−0·118	0·119	−0·117	0·120	−0·122	0·112	0·991

† Obtained by extrapolation from Pearson & Merrington's tables.

10. The test criterion discussed in the previous sections is evidently not the most sensitive possible, even for detecting a change in a pure location parameter. It would be natural to use a weighted sum $\sum_{i=1}^{k} a_i(x_i - X_i)$, the weights being chosen to make $\sum_{i=1}^{k} a_i x_i$ the best linear unbiased estimator of $\frac{1}{2}$ when the null hypothesis is valid. Gupta (1952, pp. 265–7) has calculated the coefficients for an analogous statistic, μ^*, in the case of the Normal distribution. This statistic was put forward by the author as an estimator of the mean of an unknown Normal population, given a censored sample, but it could be used to test the hypothesis we

are considering. It is therefore of some interest to compare the sensitivity of tests based on μ^* and on $\bar{x}_k$, respectively, when applied to testing the mean value of a Normal distribution with unit variance. Here we will attempt only a very rough local comparison of the two criteria, based on the 'limiting non-centrality rate' defined as

$$\frac{\partial \mathscr{E}(\text{criterion})}{\partial(\text{location parameter})} \Big/ \begin{pmatrix}\text{standard deviation}\\ \text{of criterion}\end{pmatrix}\Bigg|_{\substack{\text{null}\\ \text{hypothesis}}}.$$

For μ^* this is equal to $[\text{var}(\mu^*)]^{-\frac{1}{2}}$. For $\bar{x}_k$ we use the approximation

$$\frac{\partial \mathscr{E}(\bar{x}_k)}{\partial(\text{location parameter})} \doteqdot \frac{1}{k}\sum_{i=1}^{k}\left(Z_i - \frac{p_i q_i}{2(n+2)\,Z_i}\right),$$

where

$$Z_i = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}T_i^2}$$

and

$$\frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{T_i} e^{-\frac{1}{2}t^2}\,dt = p_i = \frac{i}{n+1}.$$

The values obtained when $n = 10$ are compared in the following table:

k	(1) $\bar{x}_k$	(2) μ^*	Ratio (2):(1)
2	2·266	0·942	0·416
4	2·722	2·056	0·755
6	2·975	2·736	0·920
8	3·129	3·050	0·975
10	3·211	3·162	0·985

The comparison is, of course, very rough and the results for $k = 10$ appear anomalous, but there does seem to be some evidence in favour of $\bar{x}_k$, at any rate for the smaller values of k. As compared with μ^*, $\bar{x}_k$ requires the additional labour of applying the probability integral transformation, but it does not require a special table of weights for its calculation, and the approximation in § 9 will often enable a rapid test of significance to be made.

11. The main purpose of the present paper has been to present formulae which it is intended to use in future work. The new applications discussed have been selected for illustrative purposes. Further subjects which it is hoped to discuss later are:

(*a*) the development of tests involving polynomials in the functions $(x_i - X_i)$ of the latter part of this paper,

(*b*) effects of non-normality on statistical procedures based on (censored) Normal distributions,

(*c*) comparison of two or more censored samples,

(*d*) extension of these methods to samples censored by fixed point(s) (Gupta's method I) and to truncated samples.

REFERENCES

CORNISH, E. A. & FISHER, R. A. (1937). *Rev. Inst. int. Statist.* **5**, 307.
GODWIN, H. J. (1949). *Ann. Math. Statist.* **20**, 279.
GUPTA, A. K. (1952). *Biometrika*, **39**, 260.
HOJO, T. (1931). *Biometrika*, **23**, 315.
PEARSON, E. S. & MERRINGTON, MAXINE P. (1951). *Biometrika*, **38**, 4.
PEARSON, K. (1931). *Biometrika*, **23**, 364.
THOMPSON, C. M. (1941). *Biometrika*, **32**, 151.

Table 1. *Cumulants of* $F(x_r)$

Note. (i) $r<s<t<v$.

(ii) For convenience we use a notation typified by the relationship

$$\kappa_{112}(F_r, F_s, F_t) = \kappa(rst^2).$$

$$p_r = \frac{r}{n+1} = 1-q_r.$$

$$\kappa(r) = p_r,$$

$$\kappa(r^2) = \frac{p_r q_r}{n+2}; \qquad \kappa(rs) = \frac{p_r q_s}{n+2},$$

$$\kappa(r^3) = \frac{2p_r q_r(q_r-p_r)}{(n+2)(n+3)}; \quad \kappa(r^2s) = \frac{2p_r q_s(q_r-p_r)}{(n+2)(n+3)},$$

$$\kappa(rs^2) = \frac{2p_r q_s(q_s-p_s)}{(n+2)(n+3)}; \quad \kappa(rst) = \frac{2p_r(q_s-p_s)q_t}{(n+2)(n+3)},$$

$$\kappa(r^4) = \frac{6p_r q_r}{(n+2)(n+3)(n+4)}\left[(q_r-p_r)^2 - \frac{n+3}{n+2}p_r q_r\right],$$

$$\kappa(r^3s) = \frac{6p_r q_s}{(n+2)(n+3)(n+4)}\left[(q_r-p_r)^2 - \frac{n+3}{n+2}p_r q_r\right],$$

$$\kappa(rs^3) = \frac{6p_r q_s}{(n+2)(n+3)(n+4)}\left[(q_s-p_s)^2 - \frac{n+3}{n+2}p_s q_s\right],$$

$$\kappa(r^2s^2) = \frac{p_r q_s}{(n+2)(n+3)(n+4)}\left[6(q_r-p_r)(q_s-p_s) + \frac{2(n+1)}{n+2}q_r p_s - \frac{4(2n+5)}{n+2}p_r q_s\right],$$

$$\kappa(r^2st) = \frac{p_r q_t}{(n+2)(n+3)(n+4)}\left[6(q_r-p_r)(q_s-p_s) + \frac{2(n+1)}{n+2}q_r p_s - \frac{4(2n+5)}{n+2}p_r q_s\right],$$

$$\kappa(rs^2t) = \frac{6p_r q_t}{(n+2)(n+3)(n+4)}\left[(q_s-p_s)^2 - \frac{n+3}{n+2}p_s q_s\right],$$

$$\kappa(rst^2) = \frac{p_r q_t}{(n+2)(n+3)(n+4)}\left[6(q_s-p_s)(q_t-p_t) + \frac{2(n+1)}{n+2}q_s p_t - \frac{4(2n+5)}{n+2}p_s q_t\right],$$

$$\kappa(rstv) = \frac{p_r q_v}{(n+2)(n+3)(n+4)}\left[6(q_s-p_s)(q_t-p_t) + \frac{2(n+1)}{n+2}q_s p_t - \frac{4(2n+5)}{n+2}p_s q_t\right].$$

Table 2. *Cumulants of ordered variables (to order* $(n+2)^{-3}$)

Note. (i) $r<s<t<v$.

(ii) For convenience we use a notation typified by the relationship

$$\kappa_{112}(x_r, x_s, x_t) = \kappa(x_r x_s x_t^2).$$

$$p_r = 1-q_r = \frac{r}{n+1}, \quad \int_{-\infty}^{X_r} dF(x) = p_r; \quad X_r' = dX_r/dF.$$

$$\mathscr{E}(x_r) = X_r + \frac{p_r q_r}{2(n+2)}X_r'' + \frac{p_r q_r}{(n+2)^2}\left[\tfrac{1}{3}(q_r-p_r)X_r''' + \tfrac{1}{8}p_r q_r X_r^{iv}\right]$$

$$+ \frac{p_r q_r}{(n+2)^3}\left[-\tfrac{1}{4}(q_r-p_r)X_r''' + \tfrac{1}{6}\{(q_r-p_r)^2 - p_r q_r\}X_r^{iv} + \tfrac{1}{6}q_r p_r(q_r-p_r)X_r^{v} + \tfrac{1}{48}p_r^2 q_r^2 X_r^{vi}\right],$$

$$\kappa(x_r^2) = \frac{p_r q_r}{n+2}X_r'^2 + \frac{p_r q_r}{(n+2)^2}\left[2(q_r-p_r)X_r'X_r'' + p_r q_r(X_r'X_r''' + \tfrac{1}{2}X_r''^2)\right]$$

$$+ \frac{p_r q_r}{(n+2)^3}\left[-2(q_r-p_r)X_r'X_r'' + \{(q_r-p_r)^2 - p_r q_r\}(2X_r'X_r''' + \tfrac{3}{4}X_r''^2)\right.$$

$$\left. + p_r q_r(q_r-p_r)(\tfrac{1}{3}X_r'X_r^{iv} + 3X_r''X_r''') + \tfrac{1}{4}p_r^2 q_r^2(X_r'X_r^{v} + 2X_r''X_r^{iv} + \tfrac{1}{3}X_r'''^2)\right],$$

Table 2 (*cont.*)

$$\kappa(x_r x_s) = \frac{p_r q_s}{n+2} X_r' X_s' + \frac{p_r q_s}{(n+2)^2}[(q_r - p_r) X_r'' X_s' + (q_s - p_s) X_r' X_s'' + \tfrac{1}{2} p_r q_r X_r''' X_s' + \tfrac{1}{2} p_s q_s X_r' X_s''' + \tfrac{1}{2} p_r q_s X_r'' X_s'']$$

$$+ \frac{p_r q_s}{(n+2)^3}[-(q_r - p_r) X_r'' X_s' - (q_s - p_s) X_r' X_s'' + \{(q_r - p_r)^2 - p_r q_r\} X_r''' X_s'$$
$$+ \{(q_s - p_s)^2 - p_s q_s\} X_r' X_s''' + \{\tfrac{3}{2}(q_r - p_r)(q_s - p_s) + \tfrac{1}{2} p_s q_r - 2 p_r q_s\} X_r'' X_s''$$
$$+ \tfrac{1}{2} p_r q_r (q_r - p_r) X_r^{iv} X_s' + \tfrac{1}{2} p_s q_s (q_s - p_s) X_r' X_s^{iv} + \{p_r q_s (q_r - p_r) + \tfrac{1}{2} p_r q_r (q_s - p_s)\} X_r''' X_s''$$
$$+ \{p_r q_s (q_s - p_s) + \tfrac{1}{2} p_s q_s (q_r - p_r)\} X_r'' X_s''' + \tfrac{1}{2} p_r^2 q_r^2 X_r'' X_s' + \tfrac{1}{2} p_s^2 q_s^2 X_r' X_s''$$
$$+ \tfrac{1}{2} p_r^2 q_r q_s X_r^{iv} X_s'' + \tfrac{1}{2} p_r p_s q_r^2 X_r'' X_s^{iv} + \tfrac{1}{18}(2 p_r^2 q_s^2 + 3 p_r q_r p_s q_s) X_r''' X_s'''],$$

$$\kappa(x_r^3) = \frac{p_r q_r}{(n+2)^2}[2(q_r - p_r) X_r'^3 + 3 p_r q_r X_r'^2 X_r'']$$

$$+ \frac{p_r q_r}{(n+2)^3}[-2(q_r - p_r) X_r'^3 + 9\{(q_r - p_r)^2 - p_r q_r\} X_r'^2 X_r'' + 3 p_r q_r (q_r - p_r)(3 X_r'^2 X_r''' + 4 X_r' X_r''^2)$$
$$+ p_r^2 q_r^2 (\tfrac{3}{2} X_r'^2 X_r^{iv} + 6 X_r' X_r'' X_r''' + X_r''^3)],$$

$$\kappa(x_r^2 x_s) = \frac{p_r q_s}{(n+2)^2}[2(q_r - p_r) X_r'^2 X_s' + 2 p_r q_r X_r' X_r'' X_s' + p_r q_s X_r'^2 X_s'']$$

$$+ \frac{p_r q_s}{(n+2)^3}[-2(q_r - p_r) X_r'^2 X_s' + 6\{(q_r - p_r)^2 - p_r q_r\} X_r' X_r'' X_s'$$
$$+ \{3(q_r - p_r)(q_s - p_s) + q_r p_s - 4 p_r q_s\} X_r'^2 X_s'' + 2 p_r q_r (q_r - p_r)(3 X_r' X_r''' X_s' + 2 X_r''^2 X_s')$$
$$+ \{2 p_r q_s (q_s - p_s) + p_s q_s (q_r - p_r)\} X_r'^2 X_s''' + \{6 p_r q_s (q_r - p_r) + 2 p_r q_r (q_s - p_s)\} X_r' X_r'' X_s''$$
$$+ p_r^2 q_r q_s (2 X_r' X_r''' X_s'' + X_r''^2 X_s') + p_r^2 q_r^2 (2 X_r'' X_r''' X_s' + X_r' X_r' X_s^{iv})$$
$$+ (p_r^2 q_s^2 + p_r q_r p_s q_s) X_r' X_r'' X_s''' + \tfrac{1}{2} p_r p_s q_r^2 X_r'^2 X_s^{iv}],$$

$$\kappa(x_r x_s^2) = \frac{p_r q_s}{(n+2)^2}[2(q_s - p_s) X_r' X_s'^2 + 2 p_s q_s X_r' X_s' X_s'' + p_r q_s X_r'' X_s'^2]$$

$$+ \frac{p_r q_s}{(n+2)^3}[-2(q_s - p_s) X_r' X_s'^2 + 6\{(q_s - p_s)^2 - p_s q_s\} X_r' X_s' X_s'' + \{3(q_r - p_r)(q_s - p_s) + p_s q_r - 4 p_r q_s\} X_r'' X_s'^2$$
$$+ 2 p_s q_s (q_s - p_s)(3 X_r' X_s' X_s''' + 2 X_r' X_s''^2) + \{2 p_r q_s (q_r - p_r) + p_r q_r (q_s - p_s)\} X_r''' X_s'^2$$
$$+ \{6 p_r q_s (q_s - p_s) + 2 p_s q_s (q_r - p_r)\} X_r'' X_s' X_s'' + p_r p_s q_s^2 (2 X_r'' X_s' X_s''' + X_r'' X_s''^2)$$
$$+ p_s^2 q_s^2 (2 X_r' X_s'' X_s''' + X_r' X_s' X_s^{iv}) + (p_s^2 q_s^2 + p_r q_r p_s q_s) X_r''' X_s' X_s'' + \tfrac{1}{2} p_s^2 q_r q_s X_r^{iv} X_s'^2],$$

$$\kappa(x_r x_s x_t) = \frac{p_r q_t}{(n+2)^2}[2(q_s - p_s) X_r' X_s' X_t' + p_r q_s X_r'' X_s' X_t' + p_s q_s X_r' X_s'' X_t' + p_s q_t X_r' X_s' X_t'']$$

$$+ \frac{p_r q_t}{(n+2)^3}[-2(q_s - p_s) X_r' X_s' X_t' + \{3(q_r - p_r)(q_s - p_s) + q_r p_s - 4 p_r q_s\} X_r'' X_s' X_t'$$
$$+ 3\{(q_s - p_s)^2 - p_s q_s\} X_r' X_s'' X_t' + \{3(q_s - p_s)(q_t - p_t) + q_s p_t - 4 p_s q_t\} X_r' X_s' X_t''$$
$$+ \{2 p_r q_s (q_r - p_r) + p_r q_r (q_s - p_s)\} X_r''' X_s' X_t' + 3 p_s q_s (q_s - p_s) X_r' X_s''' X_t'$$
$$+ \{2 p_s q_t (q_t - p_t) + p_t q_t (q_s - p_s)\} X_r' X_s' X_t''' + \{3 p_r q_s (q_s - p_s) + p_s q_s (q_r - p_r)\} X_r'' X_s'' X_t'$$
$$+ \{p_r q_s (q_t - p_t) + p_s q_t (q_r - p_r) + 2 p_r q_t (q_s - p_s)\} X_r'' X_s' X_t'' + \{p_s q_s (q_t - p_t) + 3 p_s q_t (q_s - p_s)\} X_r' X_s'' X_t''$$
$$+ \tfrac{1}{2} p_r^2 q_r q_s X_r^{iv} X_s' X_t' + \tfrac{1}{2} p_s^2 q_s^2 X_r' X_s^{iv} X_t' + \tfrac{1}{2} p_s p_t q_s^2 X_r' X_s' X_t^{iv} + \tfrac{1}{2}(p_s^2 q_s^2 + p_r q_r p_s q_s) X_r''' X_s'' X_t'$$
$$+ \tfrac{1}{2}(p_r^2 q_s q_t + p_r q_r p_s q_t) X_r''' X_s' X_t'' + p_r p_s q_r^2 X_r'' X_s''' X_t' + p_s^2 q_s q_t X_r' X_s''' X_t''$$
$$+ \tfrac{1}{2}(p_r p_s q_t^2 + p_r q_s p_t q_t) X_r'' X_s' X_t''' + \tfrac{1}{2}(p_s^2 q_t^2 + p_s q_s p_t q_t) X_r' X_s'' X_t''' + p_r p_s q_s q_t X_r'' X_s'' X_t''],$$

Table 2 (*cont.*)

$$\kappa(x_r^4) = \frac{p_r q_r}{(n+2)^3}[6\{(q_r-p_r)^2-p_r q_r\}X_r'^4+24p_r q_r(q_r-p_r)X_r'^2X_r''+4p_r^2q_r^2(X_r'^2X_r'''+3X_r''^3)],$$

$$\kappa(x_r^3x_s) = \frac{p_r q_s}{(n+2)^3}[6\{(q_r-p_r)^2-p_r q_r\}X_r'^3X_s'+18p_r q_r(q_r-p_r)X_r'^2X_r''X_s'+6p_r q_s(q_r-p_r)X_r'^3X_s''$$
$$+3p_r^2q_s^2(X_r'^2X_r'''X_s'+2X_r'X_r''^2X_s')+6p_r^2q_r q_s X_r'^2X_r''X_s''+p_r^2q_s^2X_r'^3X_s'''],$$

$$\kappa(x_r^2x_s^2) = \frac{p_r q_s}{(n+2)^3}[\{6(q_r-p_r)(q_s-p_s)-8p_r q_s+2q_r p_s\}X_r'^2X_s'^2+\{8p_r q_s(q_r-p_r)+4p_r q_r(q_s-p_s)\}X_r'X_r''X_s'^2$$
$$+\{8p_r q_s(q_s-p_s)+4p_s q_s(q_r-p_r)\}X_r'^2X_s'X_s''+2p_r^2q_r q_s(X_r'X_r'''X_s''^2+X_r''^2X_s'X_s''')$$
$$+2p_r p_s q_s^2(X_r'^2X_s'X_s'''+X_r'^2X_s''^2)+4(p_r^2q_s^2+p_r q_r p_s q_s)X_r'X_r''X_s'X_s''],$$

$$\kappa(x_r x_s^3) = \frac{p_r q_s}{(n+2)^3}[6\{(q_s-p_s)^2-p_s q_s\}X_r'X_s'^3+18p_s q_s(q_s-p_s)X_r'X_s'^2X_s''+6p_r q_s(q_s-p_s)X_r''X_s'^3$$
$$+3p_s^2q_s^2(X_r'X_s'^2X_s'''+2X_r'X_s'X_s''^2)+6p_r p_s q_s^2X_r''X_s'^2X_s''+p_s^2q_s^2X_r'''X_s'^3],$$

$$\kappa(x_r^2x_s x_t) = \frac{p_r q_t}{(n+2)^3}[\{6(q_r-p_r)(q_s-p_s)+2q_r p_s-8p_r q_s\}X_r'^2X_s'X_t'+\{8p_r q_s(q_r-p_r)+4p_r q_r(q_s-p_s)\}X_r'X_r''X_s'X_t'$$
$$+\{4p_r q_s(q_s-p_s)+2p_s q_s(q_r-p_r)\}X_r'^2X_s''X_t'+\{4p_r q_t(q_s-p_s)+2p_s q_t(q_r-p_r)\}X_r'^2X_s'X_t''$$
$$+2p_r^2q_r q_s(X_r'X_r'''X_s'X_t'+X_r''^2X_s'X_t')+p_r p_s q_t^2X_r'^2X_s'''X_t'+p_r p_s q_t^2X_r'^2X_s'X_t'''$$
$$+2(p_r^2q_s^2+p_r q_r p_s q_s)X_r'X_r''X_s''X_t'+2(p_r^2q_s q_t+p_r q_r p_s q_t)X_r'X_r''X_s'X_t''$$
$$+2p_r p_s q_s q_t X_r'^2X_s''X_t''],$$

$$\kappa(x_r x_s^2 x_t) = \frac{p_r q_t}{(n+2)^3}[6\{(q_s-p_s)^2-q_s p_s\}X_r'X_s'^2X_t'+6p_r q_s(q_s-p_s)X_r''X_s'^2X_t'+6p_s q_t(q_s-p_s)X_r'X_s'^2X_t''$$
$$+12p_s q_s(q_s-p_s)X_r'X_s'X_s''X_t'+p_r^2q_s^2X_r'''X_s'^2X_t'+p_s^2q_t^2X_r'X_s'^2X_t'''$$
$$+2p_r^2q_s^2(X_r'X_s'X_r'''X_t'+X_r'X_r''^2X_t')+4p_r p_s q_s^2X_r''X_s'X_s''X_t'$$
$$+4p_s^2q_s q_t X_r'X_s'X_s''X_t''+2p_r p_s q_s q_t X_r''X_s'^2X_t''],$$

$$\kappa(x_r x_s x_t^2) = \frac{p_r q_t}{(n+2)^3}[\{6(q_s-p_s)(q_t-p_t)+2q_s p_t-8p_s q_t\}X_r'X_s'X_t'^2+\{8p_s q_t(q_t-p_t)+4p_t q_t(q_s-p_s)\}X_r'X_s'X_t'X_t''$$
$$+\{4p_s q_t(q_s-p_s)+2p_s q_s(q_t-p_t)\}X_r''X_s'X_t'^2+\{4p_r q_t(q_s-p_s)+2p_r q_s(q_t-p_t)\}X_r'X_s'X_t'^2$$
$$+2p_s p_t q_t^2(X_r'X_s'X_t'X_t'''+X_r'X_s'X_t''^2)+p_s^2q_s q_t X_r'''X_s'X_t'^2+p_s^2q_s q_t X_r'X_s'''X_t'^2$$
$$+2(p_s^2q_t^2+p_s q_s p_t q_t)X_r'X_s''X_t'X_t''+2(p_r p_s q_t^2+p_r q_s p_t q_t)X_r''X_s'X_t'X_t''$$
$$+2p_r p_s q_s q_t X_r''X_s''X_t'^2],$$

$$\kappa(x_r x_s x_t x_v) = \frac{p_r q_v}{(n+2)^3}[\{6(q_s-p_s)(q_t-p_t)+2q_s p_t-8p_s q_t\}X_r'X_s'X_t'X_v'$$
$$+\{4p_r q_t(q_s-p_s)+2p_r q_s(q_t-p_t)\}X_r''X_s'X_t'X_v'+\{4p_s q_t(q_s-p_s)+2p_s q_s(q_t-p_t)\}X_r'X_r''X_t'X_v'$$
$$+\{4p_s q_t(q_t-p_t)+2p_t q_t(q_s-p_s)\}X_r'X_s'X_r''X_v'+\{4p_s q_v(q_t-p_t)+2p_t q_v(q_s-p_s)\}X_r'X_s'X_t'X_r''$$
$$+p_r^2q_s q_t X_r'''X_s'X_t'X_v'+p_s^2q_s q_t X_r'X_r'''X_t'X_v'+p_s p_t q_t^2X_r'X_s'X_r'''X_v'$$
$$+p_s p_t q_v^2X_r'X_s'X_t'X_r'''+2p_r p_s q_s q_t X_r''X_r''X_t'X_v'$$
$$+(p_r q_s p_t q_t+p_r p_s q_t^2)X_r''X_s'X_t'X_v'+(p_s q_s p_t q_t+p_s^2q_t^2)X_r'X_r''X_t'X_v'$$
$$+(p_r q_s p_t q_v+p_r p_s q_t q_v)X_r''X_s'X_t'X_r''+(p_s q_s p_t q_v+p_s^2q_t q_v)X_r'X_r''X_t'X_r''+2p_s p_t q_t q_v X_r'X_s'X_t'X_r''].$$

Table 3. *Sums of powers and products of the first k natural numbers*

$$\sum_{i=1}^{k} i = \frac{k(k+1)}{2},$$

$$\sum_{i=1}^{k} i^2 = \frac{k(k+1)(2k+1)}{6}, \quad \sum_{i=1}^{k-1}\sum_{j=i+1}^{k} ij = \frac{k(k-1)(k+1)(3k+2)}{24},$$

$$\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} i = \frac{k(k-1)(k+1)}{6}, \quad \sum_{i=1}^{k-1}\sum_{j=i+1}^{k} j = \frac{k(k-1)(k+1)}{3},$$

$$\sum_{i=1}^{k} i^3 = \frac{k^2(k+1)^2}{4}, \quad \sum_{i=1}^{k-1}\sum_{j=i+1}^{k} i^2 j = \frac{k(k-1)(k+1)}{120}(8k^2+5k-2),$$

$$\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} ij^2 = \frac{k(k-1)(k+1)}{120}(12k^2+15k+2),$$

$$\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} i^2 = \frac{k^2(k-1)(k+1)}{12}, \quad \sum\sum_{i<j} i^3 = \frac{k(k-1)(k+1)(3k^2-2)}{60},$$

$$\sum_{i=1}^{k-2}\sum_{j=i+1}^{k-1}\sum_{l=j+1}^{k} ijl = \frac{k^2(k+1)^2(k-1)(k-2)}{48}, \quad \sum_{i=1}^{k-2}\sum_{j=i+1}^{k-1}\sum_{l=j+1}^{k} i^2 j = \frac{k(k-1)^2(k-2)(k+1)(4k+3)}{360},$$

$$\sum\sum\sum_{i<j<l} ij^2 = \frac{k(k-1)(k-2)(k+1)^2(2k-1)}{120}, \quad \sum\sum\sum_{i<j<l} i^2 l = \frac{k(k-1)(k-2)(k+1)(5k^2+k-3)}{360},$$

$$\sum\sum\sum_{i<j<l} il^2 = \frac{k(k-1)(k-2)(k+1)(10k^2+14k+3)}{360},$$

$$\sum\sum\sum_{i<j<l} ij = \frac{k(k-1)(k-2)(3k+1)(k+1)}{120}, \quad \sum\sum\sum_{i<j<l} il = \frac{k(k-1)(k-2)(k+1)(4k+3)}{120},$$

$$\sum_{i=1}^{k} i^4 = \frac{k(k+1)(6k^3+9k^2+k-1)}{30},$$

$$\sum_{i=1}^{k} i^5 = \frac{k^2(k+1)^2}{12}(2k^2+2k-1),$$

$$\sum_{i=1}^{k} i^6 = \frac{k(k+1)}{42}(6k^5+15k^4+6k^3-6k^2-k+1),$$

$$\sum\sum_{i<j} i^2 j^2 = \frac{k(k+1)(k-1)(20k^3+24k^2-5k-6)}{360},$$

$$\sum\sum\sum_{i<j<l} i = \frac{k(k-1)(k-2)(k+1)}{24}, \quad \sum\sum\sum_{i<j<l} i^2 = \frac{k(k-1)(k-2)(2k-1)(k+1)}{120},$$

$$\sum\sum\sum_{i<j<l} i^2 jl = \frac{k(k-1)(k-2)(k+1)}{5040}(48k^3+19k^2-61k-24),$$

$$\sum\sum\sum_{i<j<l} ij^2 l = \frac{k(k-1)(k-2)(k+1)}{5040}(72k^3+81k^2-39k-36),$$

$$\sum\sum\sum_{i<j<l} ijl^2 = \frac{k(k-1)(k-2)(k+1)}{5040}(90k^3+145k^2+23k-24),$$

$$\sum\sum\sum\sum_{i<j<l<t} i = \frac{k(k-1)(k-2)(k-3)(k+1)}{120},$$

$$\sum\sum\sum\sum_{i<j<l<t} ij = \frac{k^2(k-1)(k-2)(k-3)(k+1)}{240},$$

Table 3 (*cont.*)

$$\sum_{i<j<l<t}\sum il = \frac{k(k-1)(k-2)(k-3)(k+1)(2k+1)}{360},$$

$$\sum_{i<j<l<t}\sum it = \frac{k(k-1)(k-2)(k-3)(k+1)(5k+4)}{720},$$

$$\sum_{i<j<l<t}\sum ijl = \frac{k(k-1)(k-2)(k+1)(k-3)(15k^2+5k-8)}{5040},$$

$$\sum_{i<j<l<t}\sum ijt = \frac{k(k-1)(k-2)(k-3)(k+1)(18k^2+13k-4)}{5040},$$

$$\sum_{i<j<l<t}\sum ilt = \frac{k(k-1)(k-2)(k-3)(k+1)(24k^2+29k+4)}{5040},$$

$$\sum_{i<j<l<t}\sum ijlt = \frac{k(k-1)(k-2)(k-3)(k+1)(15k^3+15k^2-10k-8)}{5760}.$$

SIMPLIFIED DECISION FUNCTIONS

By G. A. BARNARD
Imperial College

The recent work of Wald (1950) and Lindley (1953) in the theory of statistical decision functions is of the utmost importance to practising statisticians, but it seems at present rather inaccessible to these latter because of the formidable mathematical apparatus with which it appears to be connected. Our object in this paper is to give a reasonably simple proof of a result closely related to that proved by Lindley, and indirectly related to Wald's central theorems. Our result can also be regarded as a generalization of Neyman & Pearson's fundamental lemma (1933), and it thus forms a bridge from their work to the later work of Wald and Lindley. We arrive at our result by means of two lemmas, one geometrical and one statistical. The geometrical lemma is a purely mathematical result, and its proof would not have been given in this paper had it been longer or more complicated than it is. The reader who wishes to concentrate on the statistical argument can omit the section devoted to this lemma. The statistical lemma, on the other hand, is of some interest in itself as giving a direct frequency interpretation of a likelihood ratio.

Our result on decision functions is essentially a slight extension of that first derived by Lindley, but for simplicity we have restricted ourselves to the case where only two decisions are possible. In this respect our results could be extended at the expense of tedious but inessential complication. Our methods of proof, too, have been derived from those of Lindley, by omission of these complications, by use of our geometrical lemma, and by simplification caused by our adoption from the beginning of the notion of a randomized decision function. Lindley went to some trouble to avoid randomized decisions as much as possible (they cannot be altogether avoided), in what now seems to the present writer a misguided endeavour. If we confuse the decision problem with the inference problem, randomization appears as a thing to be avoided. It *is* unreasonable that the *conclusion* to be drawn from a set of data should depend on the totally irrelevant throw of a coin; but it *is not* unreasonable that the *action* to be taken on the basis of certain facts should be decided by the throw of a coin, if the pros and cons are very evenly balanced.

A formulation of probability theory

In any work on general probability theory one is harassed by the difficulty that everything has to be said twice, once for discrete and once for continuous variates. One might say, three times, if one adds the multivariate case. It has for some time been customary to bring the first two cases together by using the cumulative distribution function instead of the probability (density) function; but this, besides being of no particular help in the multivariate case, is excessively awkward when problems of inference or decision are involved. We therefore propose another way out, by a reinterpretation of existing symbolism.

We propose to take as our general expression for the probability of a proposition A depending on a single trial

$$\Pr\{A \mid \phi\} = \int A(x)\,\phi(x)\,dx,$$

where x is a variable ranging over the set of values possible for the variate X under discussion, $\phi(x)$ is a *continuous* function of x, called the probability function, and $A(x)$ is a function of x defined by saying that $A(x)$ is 1 when A is true and is 0 when A is false. The meaning of the pair of symbols

$$\int \dots dx$$

depends on the nature of the range of x. If x is a single real variable, these symbols have their ordinary meaning (the range being from $-\infty$ to $+\infty$). If x is restricted to being a positive whole number, they are to be taken as meaning

$$\Sigma \dots,$$

while if, for example, x represents a point (x_1, x_2, x_3) in continuous three-dimensional space, they are to be taken as meaning

$$\iiint \dots dx_1 dx_2 dx_3.$$

If the truth of a proposition B depends on the results of two (independent) observations on X, its 'characteristic function' $B(x, y)$ will be a function of two variables x, y, corresponding to the results of the two trials; $B(x, y)$ will again, by definition, take the value 1 when B is true and 0 when B is false. For the probability of such a proposition we shall need to interpret

$$\iint \dots dx dy.$$

We do this in the obvious manner. If, for example, x is the point (x_1, x_2, x_3), while y is the point (y_1, y_2, y_3), then we take

$$\iint \dots dx dy \quad \text{to mean} \quad \iiiint\!\!\iiint \dots dx_1 dx_2 dx_3 dy_1 dy_2 dy_3,$$

and similarly in other cases. We then have

$$\Pr\{B \mid \phi\} = \iint B(x, y)\, \phi(x)\, \phi(y)\, dx dy,$$

since, by the multiplication law of probability, the probability function for two independent trials (x, y) is the product $\phi(x)\,\phi(y)$ of the probability functions for each trial separately.

For the mathematically minded reader we can sum this up by saying that, in accordance with modern practice, we are taking

$$\int \dots dx$$

to represent a Stieltjes integral with respect to a measure supposed defined in the space of X. We depart from modern practice among probabilists by insisting that in all cases arising in practice this space may be taken to be a topological space, and that $\phi(x)$ can be taken continuous in this topology, while the given measure is defined and non-null for all non-empty open sets. The consideration of more general cases, we would argue, belongs to pure mathematics, not to statistics as a branch of applied mathematics. Perhaps the principal logical advantage of our approach is based on the fact that for us the probability function is defined everywhere, whereas for the more general probabilistic approach the probability function is defined only almost everywhere.

A GEOMETRICAL LEMMA

In this section (not in other sections) $x = (x_1, x_2, \ldots, x_n)$ will be taken to represent a point in n-dimensional vector space. If x, y are points in this space, and α and β are real numbers, the point $\alpha x + \beta y$ is that with co-ordinates $(\alpha x_1 + \beta y_1, \alpha x_2 + \beta y_2, \ldots, \alpha x_n + \beta y_n)$, and the 'inner product' $x.y$ denotes the real number

$$x.y = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n.$$

We call the positive aliquant (cf. quadrant, octant) the set of all points none of whose co-ordinates are negative, and the negative aliquant the set of all points none of whose co-ordinates are positive. These aliquants are not supposed to contain the origin $(0, 0, \ldots, 0)$. A set of points C is said to be convex if, whenever it contains the points x and y, it contains all points on the line segment joining x to y, i.e. all points $\theta x + (1-\theta) y$ with $0 \leqslant \theta \leqslant 1$.

A set C is said to 'have the property P' if (a) C is convex, (b) no point of C is interior to the negative aliquant. We now prove the

LEMMA. *If C has the property P, there exists a point u in the positive aliquant such that*

$$u.x \geqslant 0$$

for all x in C.

Proof. We use induction on the dimension number n. Assume first that $n = 2$ and consider the points $(\cos\alpha, \sin\alpha)$ on the unit circle in the first quadrant, $0 \leqslant \alpha \leqslant \frac{1}{2}\pi$. We put α into class L if there is an x in C in the second quadrant such that $x_1 \cos\alpha + x_2 \sin\alpha < 0$, and we put α into class U if there is an x in C in the fourth quadrant such that $x_1 \cos\alpha + x_2 \sin\alpha < 0$. Then if α is in L and β is in U, we must have $\alpha < \beta$. For otherwise, if x is a point of C which shows α to be in L, and y is a point of C which shows β to be in U, the segment joining x to y passes through the interior of the third quadrant, contra hypothesis. Further, since $x_1 \cos\alpha + x_2 \sin\alpha$ is a continuous function of α, the sets L and U are both open, i.e. L has no greatest member and U has no least. It follows that there must be at least one α in neither class, and for this α we have

$$x_1 \cos\alpha + x_2 \sin\alpha \geqslant 0$$

for all x in C in the second, fourth and first quadrants, i.e. for all x in C.

Now assume the lemma true for dimensions $2, 3, \ldots, (n-1)$, and suppose C to be a set in n dimensions. We define an $(n-1)$-dimensional set C^- by saying that $(x_1, x_2, \ldots, x_{n-1})$ belongs to C^- if and only if, for some $x_n \leqslant 0$, $(x_1, x_2, \ldots, x_{n-1}, x_n)$ belongs to C. It is easily seen that C^- has the property P, so that there exist, by the inductive hypothesis, $v_1, v_2, \ldots, v_{n-1}$, non-negative and not all zero, such that

$$v_1 x_1 + v_2 x_2 + \ldots + v_{n-1} x_{n-1} \geqslant 0$$

for all $(x_1, x_2, \ldots, x_{n-1})$ in C^-. Thus if v denotes the point $(v_1, v_2, \ldots, v_{n-1}, 0)$, we have for all points x in C, either $x_n > 0$, or $v.x \geqslant 0$, or both. Thus the two-dimensional set of points $(v.x, x_n)$, where x ranges over C, satisfies condition (b). It is easily seen to be convex, and so, by the first part, there exist non-negative w_1, w_2, not both 0, such that for all x in C,

$$w_1 (v.x) + w_2 x_n \geqslant 0.$$

Thus we can here take u to be the point $(w_1 v_1, \ldots, w_1 v_{n-1}, w_2)$ and the lemma is proved.

A STATISTICAL LEMMA

Now let us consider situations where we are able to make independent observations on a variate X whose probability function ϕ is known to be one of a finite number of functions $\phi_0, \phi_1, \ldots, \phi_k$. We wish to decide whether or not to take a certain action A which is desirable if $\phi = \phi_0$, but which is undesirable if $\phi = \phi_i$ for $i = 1, 2, \ldots, k$. In Neyman & Pearson's terminology we might be said to be 'testing the hypothesis' $\phi = \phi_0$ against the set of alternatives $\phi = \phi_i$, and action A would then correspond to 'acceptance' of the hypothesis being tested.

Any rule for making the decision will consist of two parts:

(i) a rule for deciding when to stop taking observations,

(ii) a rule for deciding whether or not to take action A, given that observations have ceased.

We call (i) the sampling rule, and (ii) the (final) decision rule. In making decisions in practical life we often find the pros and cons so evenly balanced that the only way of arriving at a definite decision is to 'toss for it'. This means that we allow our decision to depend not only on the relevant data, but also on the result of an irrelevant random experiment. We allow for this possibility in our theory, partly because of its occurrence in practical cases, but mainly for its mathematical convenience. With this in mind we can say that a sampling rule will be given by a sequence of functions $s_1, s_2, \ldots, s_r, \ldots, s_n$, such that

(i) s_r is a function of r variables $(x_1, x_2, \ldots, x_r)$ such that $0 \leqslant s_r(x_1, x_2, \ldots, x_r) \leqslant 1$ for all x_i and $r = 1, 2, \ldots, n$,

(ii) $s_n(x_1, x_2, \ldots, x_n) = 1$ for all x_i.

$s_r(x_1, x_2, \ldots, x_r)$ is to be interpreted as the probability, given that r observations have been made, with results $x_1, x_2, \ldots, x_r$, that observations will then cease. Thus condition (ii) implies that observations will certainly cease at the nth observation, if not before, so that n is an upper bound to the sample size, which exists in all practical cases. Our argument is made slightly simpler by assuming that n is finite.

The probability that the total number of observations taken is r is the probability function $\phi(x_1)\,\phi(x_2) \ldots \phi(x_r)$ for the sequence $x_1, x_2, \ldots, x_r$, multiplied by the probability of stopping observations at this point, integrated over all possible sequences of length r, i.e.

$$\int\int \ldots \int (1-s_1)\,(1-s_2) \ldots (1-s_{r-1})\, s_r \phi(x_1) \ldots \phi(x_r)\, dx_1 \ldots dx_r.$$

Turning now to final decisions, we formally define:

A final decision rule is a sequence of functions $d_1, d_2, \ldots, d_r, \ldots, d_n$, such that for all sequences of values of x, $0 \leqslant d_r(x_1, x_2, \ldots, x_r) \leqslant 1$ for $r = 1, 2, \ldots, n$.

Such a rule has the interpretation that, if the values $x_1, x_2, \ldots, x_r$ have been observed for X, and if the sampling rule has given the result that observations are to cease, then a random experiment is to be performed in which the probability of 'success' is $d_r(x_1, x_2, \ldots, x_r)$, and if 'success' is obtained in the experiment, then action A is to be taken, and not otherwise. We then have, for the probability of action A, the total probability of taking action A after one observation, after two observations, ..., after n observations:

$$\Pr\{A \mid \phi\} = \int d_1 s_1 \phi(x_1)\, dx_1 + \int\int d_2 (1-s_1)\, s_2 \phi(x_1)\, \phi(x_2)\, dx_1 dx_2 + \ldots$$

$$+ \int\int \ldots \int d_n (1-s_1)\,(1-s_2) \ldots (1-s_{n-1})\, s_n \phi(x_1)\, \phi(x_2) \ldots \phi(x_n)\, dx_1 dx_2 \ldots dx_n.$$

We now concentrate attention on the 'hypothesis being tested', ϕ_0, and one of the alternatives ϕ_i. Following Fisher we define the *likelihood ratio* for the sequence of observations $x_1, x_2, \ldots, x_r$ to be

$$l_{ir}(x_1, x_2, \ldots, x_r) = \phi_i(x_1)\,\phi_i(x_2) \ldots \phi_i(x_r)/\phi_0(x_1)\,\phi_0(x_2) \ldots \phi_0(x_r),$$

and adapting Neyman & Pearson we define the

$$\text{Odds of error} = \frac{\text{Probability of wrongly taking action } A, \text{ when } \phi = \phi_i}{\text{Probability of rightly taking action } A, \text{ when } \phi = \phi_0}$$

$$= \Pr\{A \mid \phi_i\}/\Pr\{A \mid \phi_0\};$$

or, in the customary notation of 'errors of the first and second kinds', $\beta/(1-\alpha)$. With these definitions our statistical lemma can be stated:

Odds of error = Conditional mean value of likelihood ratio, given action A correctly taken.

The proof is immediate. Both sides can be seen to be equal to

$$\frac{\sum_{r=1}^{n} \int\int \ldots \int l_{ir} d_r (1-s_1)(1-s_2) \ldots (1-s_{r-1}) s_r \phi_0(x_1) \ldots \phi_0(x_r)\, dx_1 \ldots dx_r}{\sum_{r=1}^{n} \int\int \ldots \int d_r (1-s_1)(1-s_2) \ldots (1-s_{r-1}) s_r \phi_0(x_1) \ldots \phi_0(x_r)\, dx_1 \ldots dx_r},$$

since $$l_{ir}\phi_0(x_1) \ldots \phi_0(x_r) = \phi_i(x_1) \ldots \phi_i(x_r).$$

Although it is so simple, this lemma has many applications. For example, if the decision rule is such that the likelihood ratio is always less than a fixed number L whenever action A is taken, the conditional mean value must be less than L, and so must the odds of error be. This is a slight sharpening of an inequality due to C. A. B. Smith. It provides a justification for 'equating' the likelihood ratio to the odds of error. If action A consists in asserting that ϕ_0 is 'moderately likely', and we say this when the likelihood ratio lies between 1/20 and 1/10, then in the long run the ratio of the relative frequency of saying ϕ_0 is moderately likely when it is false (and ϕ_i is true), to the relative frequency of saying ϕ_0 is moderately likely when it is true, lies between 1/20 and 1/10. The special interest of this frequency interpretation of the likelihood ratio lies in the fact that, like the likelihood ratio itself, it is independent of the 'reference set'.

If we now turn from considering one alternative hypothesis ϕ_i to the whole set of alternatives $i = 1, 2, \ldots, k$, it is convenient to introduce a set of non-negative weights w_i, such that $\sum_i w_i = 1$, and then to define the weighted odds of error as

$$\sum_i w_i \Pr\{A \mid \phi_i\}/\Pr\{A \mid \phi_0\},$$

and the weighted likelihood ratio as $l_r = \sum_i w_i l_{ir}$.

Then since the mean value of a weighted mean is the weighted mean of the mean values, we have that the weighted odds of error will be equal to the conditional mean value of the weighted likelihood ratio. Then, as a direct application of this result, we may mention that if we arrange the decision rule so that action A is taken only when the weighted likelihood ratio is less than a fixed number L, the weighted odds of error will then be less than L. In particular, since a weighted mean of a set of quantities lies between the highest and lowest of these quantities, it follows that if the probabilities of error are all equal, they will all be less than L.

Admissible decision rules and Bayes rules

If we now adopt Neyman & Pearson's terminology, and allow the action A to correspond to acceptance of the hypothesis ϕ_0, then for any given decision rule the graph of the odds of error for ϕ_i against the suffix i will represent what they call the power curve of the decision rule—inverted, and scaled up in the ratio $1:\Pr\{A \mid \phi_0\} = 1:(1-\alpha)$, in the usual notation. In this section we shall follow Neyman & Pearson in considering that the sampling rule is fixed, and the probability of rightly accepting ϕ_0, $\Pr\{A \mid \phi_0\}$, is also fixed at a value ρ, and then we compare different decision rules on the basis of their power curves.

If the power curve for a decision rule D never lies below the power curve for a decision rule D', and sometimes lies above it, we say that the decision rule D is uniformly more powerful than D', or, in Lindley's terms, D is uniformly 'better' than D'. And we say that D is *admissible* if no rule exists which is uniformly better, or more powerful, than it. In other words, D is admissible if there is no rule D' for which, simultaneously

$$\Pr\{A \mid D', \phi_0\} = \Pr\{A \mid D, \phi_0\} = \rho,$$

and for all i, $$\Pr\{A \mid D', \phi_i\} \leqslant \Pr\{A \mid D, \phi_i\},$$

while for some i, $$\Pr\{A \mid D', \phi_i\} < \Pr\{A \mid D, \phi_i\}.$$

If D is inadmissible, then a uniformly more powerful rule D' exists, and the only possible reason for using D rather than D' to decide one's course of action would be that D involved less computation than D', or something similar. On the other hand, if D is admissible no conclusive argument based solely on the frequency of errors can be brought against its use. We now proceed to characterize the class of all admissible rules . If this class turns out to have only one member, this rule will undoubtedly be the one to use (apart from considerations of ease of computation etc.); it will in fact be the uniformly most powerful rule. In general, of course, the class of all admissible rules will have more than one member.

Let D be a fixed admissible final decision rule, and let D' range over all possible final decision rules such that $$\Pr\{A \mid D, \phi_0\} = \Pr\{A \mid D', \phi_0\} = \rho,$$

and consider the set C of points x in k-dimensional vector space such that the ith co-ordinate x_i of x is

$$x_i = \Pr\{A \mid D', \phi_i\} - \Pr\{A \mid D, \phi_i\}.$$

Then C has the property P of our geometrical lemma. For if D', corresponding to the point x in C, has typical function d'_r, and D'', corresponding to the point x' in C, has typical function d''_r, then the point $\theta x + (1-\theta)x'$, for $0 \leqslant \theta \leqslant 1$, corresponds to the decision rule $\theta D' + (1-\theta)D''$, with typical function $\theta d'_r + (1-\theta)d''_r$, and for this decision rule we have

$$\begin{aligned}\Pr\{A \mid \theta D' + (1-\theta)D'', \phi_0\} &= \theta \Pr\{A \mid D', \phi_0\} + (1-\theta)\Pr\{A \mid D'', \phi_0\} \\ &= \theta\rho + (1-\theta)\rho = \rho.\end{aligned}$$

Hence C is convex. It has no points in the negative aliquant because D is admissible. And so, from the geometrical lemma, it follows that there is a point u in the positive aliquant for which

$$u.x \geqslant 0$$

for all x in C. If we take $w_i = u_i/\Sigma u_i$ this means that non-negative weights w_i exist, with $\sum_i w_i = 1$, such that for all D' of the class considered

$$\sum_i w_i \Pr\{A \mid D, \phi_i\} \leqslant \sum_i w_i \Pr\{A \mid D', \phi_i\} \quad (i = 1, 2, \ldots, k),$$

or, dividing by ρ, we see that D minimizes the weighted odds of error among all rules giving the same probability of rightly taking action A.

We shall call a rule which minimizes the weighted odds of error, for some choice w_i of weights, a Bayes rule. We have just seen that all admissible rules are Bayes rules. We now proceed to justify the term 'Bayes rule', and to show that in a sense the converse is true—'almost all' Bayes rules are admissible.

First we notice that minimizing the weighted odds of error is, by the statistical lemma, equivalent to minimizing the conditional mean value of the weighted likelihood ratio l_r. And since we are considering the sampling rule as fixed, the likelihood ratio l_r can be regarded as the value of a variate Λ having a definite distribution, with cumulative function

$$\Pr\{\Lambda \leqslant \lambda \mid \phi_0\} = F(\lambda).$$

We evidently minimize the conditional mean value of Λ by picking out the smallest values first, i.e. by choosing $d_r(x_1, x_2, \ldots, x_r)$ so that

$$\begin{aligned} d_r &= 1 \text{ when } \Lambda \leqslant \lambda(\rho) \\ &= 0 \text{ otherwise,} \end{aligned}$$

where $\lambda(\rho)$ is such that $F(\lambda(\rho)) = \rho$. If the distribution of Λ is not continuous, such a $\lambda(\rho)$ may not exist; if so, there will be a λ such that

$$\Pr\{\Lambda < \lambda \mid \phi_0\} < \rho,$$

while

$$\Pr\{\Lambda \leqslant \lambda \mid \phi_0\} > \rho.$$

In such a case we choose θ so that

$$\Pr\{\Lambda < \lambda \mid \phi_0\} + \theta \Pr\{\Lambda = \lambda \mid \phi_0\} = \rho,$$

and then put

$$\begin{aligned} d_r &= 1 \text{ when } \Lambda < \lambda \\ &= \theta \text{ when } \Lambda = \lambda \\ &= 0 \text{ otherwise.} \end{aligned}$$

Alternatively, when $\Lambda = \lambda$ we may give d_r any set of values whose conditional mean is θ, instead of putting $d_r = \theta$.

Now since w_i is non-negative and $\sum_i w_i = 1$, the weights w_i could be thought of as corresponding to prior probabilities, w_i being the prior probability of ϕ_i, given that one of the alternatives ϕ_i is true. Then if p is the prior probability of ϕ_0, and $q = 1-p$, we have, by Bayes's theorem, when the results $x_1, x_2, \ldots, x_r$ have been observed,

$$\begin{aligned} \text{Posterior odds against } \phi_0 &= \sum_i q w_i \phi_i(x_1) \ldots \phi_i(x_r) / p\phi_0(x_1) \ldots \phi_0(x_r) \\ &= q c_r / p. \end{aligned}$$

Thus it will be odds on that ϕ_0 is true if $c_r < p/q$

and it will be evens if $c_r = p/q$.

Thus if we had a person using Bayes's theorem, with this assessment of the prior probabilities, and with the rule that he was prepared to take action A whenever the odds favoured ϕ_0, and to 'toss for it' when the odds were even, such a person would use the same rule as that obtained by minimizing the weighted odds of error, provided p/q were taken equal to ρ, and the 'toss' were such that the probability of 'success' was the θ given in the preceding paragraph. It would be impossible, on the basis of an objective study of his behaviour in

situations such as we are considering, to differentiate between such a person and a person who acted on the rule of the preceding paragraph. This is why we call such rules 'Bayes rules'.

We should perhaps observe at this point that in general the specification of a Bayes rule will involve the specification of the quantitative losses expected to be incurred from taking wrong decisions, in addition to the prior probabilities. It happens in our case that this is unnecessary because we are considering only a two-way decision—to take action A, or not. Thus in extending our treatment to cases of multiple decisions we would have to introduce this slight extra complication. The general Bayes principle will be to take action A when the expected gain from doing so outweighs the expected loss.

Now in our formulation of the decision problem, whether or not to take action A, we have said nothing about how the true ϕ comes to be determined. It is therefore consistent with our assumptions (though it does not follow from them) to assume further that the true ϕ is determined by a random procedure in which the probability that $\phi = \phi_i$ is qw_i, while the probability that $\phi = \phi_0$ is p. With this further assumption, it is clear that the Bayes rule given above is best possible, subject to the condition that $\Pr\{A \mid \phi_0\} = \rho$. Any other rule would mean that we could find two situations such that the posterior probability of ϕ_0 in the first situation was less than it was in the second situation, and yet we would take action A in the first situation but not in the second; we would clearly then improve our chances of acting rightly if we interchanged our actions in the two cases.

Now this argument shows that every Bayes rule with non-zero weights w_i is admissible. For if there were a uniformly more powerful rule we could, by using it, in the case where our further assumption is true, reduce the frequency with which action A was wrongly taken, without decreasing the frequency with which action A was rightly taken; and we have just seen this to be impossible. If some of the weights w_i are zero, the argument is not conclusive. For it might happen that the uniformly more powerful rule reduced the probabilities of error only for those ϕ_i for which the weights w_i were zero. We would then not improve on the Bayes rule in those cases where the further assumption was true; but we might improve on the Bayes rule in other cases. What does follow in this case is that there can be no rule which gives smaller risks of error in all those cases ϕ_i for which the weights w_i are non-zero. If we say that D is weakly admissible if no D' exists such that

$$\Pr\{A \mid D', \phi_0\} = \Pr\{A \mid D, \phi_0\},$$

while

$$\Pr\{A \mid D', \phi_i\} < \Pr\{A \mid D, \phi_i\} \quad \text{for all } i = 1, 2, \ldots, k,$$

then our argument shows that every Bayes rule is weakly admissible.

Thus, to sum up, we have shown that every admissible rule is a Bayes rule, and every Bayes rule is weakly admissible, while every Bayes rule with non-zero weights is admissible. By going deeper we could clearly prove more. For example, under suitable continuity restrictions we could show that any Bayes rule with some weights zero could be obtained as a limit from a sequence of Bayes rules with non-zero weights, and then such a Bayes rule would be admissible, not merely weakly admissible. More simply, we may observe that any Bayes rule is admissible for the restricted set of alternatives for which the weights are non-zero. So that in any case we may say that, in so far as we adopt the approach of Neyman & Pearson, as carried further by Wald and Lindley, we are led to consider the class of all Bayes solutions to the decision problem as containing all the solutions worth consideration, and as containing no solution wholly objectionable.

ADMISSIBLE SAMPLING RULES

Before concluding we would like to indicate some of the points involved in extending our considerations to the question of selecting 'admissible sampling rules'. We therefore relax the restriction so far made that the sampling rule is fixed. We now consider sampling rules which are, in some sense, 'equivalent' in respect of admissible final decisions. It seems reasonable to assume that the weights w_i to be used in making the final decision are fixed for us, since their choice will be conditioned mainly by the relative importance of the various sorts of mistake which might be involved in wrongly taking action A, or in wrongly failing to take it. Thus the likelihood ratio Λ to be used in making our final decision will be a given function of the observations. And from the preceding considerations it seems reasonable to say that one sampling rule will be equivalent to another in respect of final decisions if the associated cumulative distribution function of Λ is the same for both. This will imply that, for every choice of $\Pr\{A \mid \phi_0\} = \rho$, the minimum value of the weighted odds of error will be the same for all sampling rules considered; conversely, this condition implies the former one. We thus consider only sampling schemes for which the cumulative distribution function of Λ is fixed as $F(\lambda)$. These sampling schemes will differ amongst themselves in the numbers of observations called for when they are used. We shall want the number of observations to be as small as possible, in some sense.

In general, the number of observations called for by a sampling rule will be a variate whose distribution depends on which of the possible hypotheses happens to be the true one; but in order to keep things simple we shall consider only one probability function ϕ. Then if the typical function of the sampling rule S is $s_r(x_1, x_2, \ldots, x_r)$, and we define

$$t_r = (1-s_1)(1-s_2)\ldots(1-s_r),$$

then the probability that the sampling rule S calls for a sample larger than r is

$$\Pr\{N > r \mid S\} = \iint\ldots\int t_r \phi(x_1)\ldots\phi(x_r)\,dx_1\ldots dx_r.$$

We now define an admissible sampling rule to be a sampling rule S which is such that no S' exists, equivalent to S in respect of final decisions, and such that

$$\Pr\{N > r \mid S'\} \leqslant \Pr\{N > r \mid S\} \quad \text{for all } r = 1, 2, \ldots, (n-1)$$

while for some r

$$\Pr\{N > r \mid S'\} < \Pr\{N > r \mid S\}.$$

Now suppose S is admissible, and consider the n-dimensional set C of points u such that the rth co-ordinate of u is

$$u_r = \Pr\{N > r \mid S'\} - \Pr\{N > r \mid S\},$$

where S' ranges over all sampling rules, admissible or not, which are equivalent to S in respect of final decisions. Then C has the property P of our geometrical lemma. It is easy to see that no point of C is interior to the negative aliquant. We have to show that C is convex.

Let us suppose that S', with typical function s'_r, and S'', with typical function s''_r, both satisfy the conditions laid down. We define

$$t'_r = (1-s'_1)(1-s'_2)\ldots(1-s'_r), \quad p'_r = s'_r t'_{r-1}$$

and t_r'' and p_r'' are defined similarly. We then define another sampling rule Σ by the conditions

$$\sigma_1 = (s_1' + s_1'')/2, \quad \sigma_r = (s_r' t_{r-1}' + s_r'' t_{r-1}'')/(t_{r-1}' + t_{r-1}'').$$

Since σ_r is a weighted mean of s_r' and s_r'' it lies between them, and hence it lies between 0 and 1, and so it represents the typical function of a possible sampling rule. We also have

$$\tau_r = (1-\sigma_1)(1-\sigma_2)\dots(1-\sigma_r) = \{(t_r' + t_r'')/(t_{r-1}' + t_{r-1}'')\}\tau_{r-1}$$
$$= (t_r' + t_r'')/2,$$

and hence $$\pi_r = \sigma_r \tau_{r-1} = (p_r' + p_r'')/2.$$

Thus the probability that a final decision will be taken on the basis of the sequence of observations $(x_1, x_2, \dots, x_r)$ when the sampling rule Σ is used is half-way between that when the sampling rule S' is used and that when the sampling rule S'' is used. It follows that since $F(\lambda)$ is the cumulative probability function for Λ when either S' or S'' is used, $F(\lambda)$ is still the cumulative probability function for λ when Σ is used. Thus Σ is equivalent to S in respect of final decisions, and corresponds to a point ν in C for which the rth co-ordinate

$$\nu_r = (u_r' + u_r'')/2,$$

where u_r' and u_r'' are the rth co-ordinates of the points in C corresponding to S' and S''. Thus when C contains two points it contains the midpoint of the line joining them, and so is convex in a slightly weaker sense than the usual one. It can easily be seen that our proof of the geometrical lemma remains valid for this weaker sense of convexity.

It now follows that if S is an admissible sampling rule it must be such as to minimize a weighted sum

$$\sum_r C_r \Pr\{N > r \mid \phi\},$$

with non-negative C_r. These C_r can clearly be interpreted as the extra cost involved in taking more than r observations given that we have already taken r. If this is done, then S will be a rule which minimizes the expected cost of the observations. Such a rule is called a Bayes sampling rule, and this terminology can be justified, and a converse proved, as with the final decision rules.

We cannot proceed further, to consider the simultaneous changing of the final decision rule and the sampling rule without, as it were, establishing a rate of exchange which decides how we are to balance a gain in the reliability of our decision against an increase in the number of observations. Wald's completely pragmatic approach seems to be the only way of doing this.

Conclusion—uniformly most powerful rules and inference

Reverting now to the case where the sampling rule is fixed, it may turn out in particular cases that there is only one admissible decision rule. This will be so if the sets of observations for which $\Lambda < \lambda(\rho)$ are independent of the choice of weights w_i. Now this is the classical condition of Neyman & Pearson for the existence of a uniformly most powerful test. And as we have already remarked if there is only one admissible rule, it must be uniformly most powerful.

In other cases, where no uniformly most powerful rule exists, the choice among the admissible rules must be based on features of the problem not so far taken into account—features such as would lead, for example, to at least an approximate specification of a prior

probability distribution. In the absence of such further information the decision problem has an essentially arbitrary character, which sharply differentiates it from the inference problem. An arbitrary inference cannot be valid in any objective sense; thus we are led to follow Fisher, in basing our inferences on likelihood, which is uniquely determined by the observations.

There is here an analogy with diagnosis and treatment in medical practice. The inference from the data is the diagnosis; the decision adopted is the treatment. The diagnosis does not uniquely determine the treatment, in deciding on which the doctor will have regard to other features of the case besides the patient's specific illness.

A tendency exists in the literature to play down the arbitrariness of the decision problem, sometimes by the introduction of arbitrary principles like Bayes's axiom, the Neyman conditions of 'unbiasedness' and 'similarity', or the minimax principle, to name some in historical order. It seems to the present writer that this tendency arises from a mistaken attempt to identify the decision problem with the inference problem. An inference from data should be unique, though it may be uncertain. But to try to give rules for unique decisions, in the absence of elements essential to a unique solution is like trying to design a bridge without knowing the strength of the materials available—only very crude approximations will be possible.

REFERENCES

LINDLEY, D. V. (1953). *J. R. Statist. Soc.* B, **14**, 43.
NEYMAN, J. & PEARSON, E. S. (1933). *Proc. Camb. Phil. Soc.*, **29**, 492.
WALD, A. (1950). *Statistical Decision Functions.* New York: John Wiley and Sons, Inc.

MISCELLANEA

Sequential tests for binomial and exponential populations

By F. J. ANSCOMBE and E. S. PAGE, *Statistical Laboratory, Cambridge*

Consider a series of independent trials in each of which a certain event has probability p of occurring. If the trials take place at equally spaced time-instants, so that the time elapsing from one trial to the next is always τ, and if we consider the limit $\tau \to 0$ and $p \to 0$ with p/τ constant, say $p/\tau = \lambda$, then in the limit the occurrences of the event form a Poisson process with mean rate λ per unit time. Let t_1 denote the time elapsing from an arbitrary origin to the first following occurrence of the event in the Poisson process, and let $t_r\,(r \geqslant 2)$ denote the time elapsing from the $(r-1)$th to the rth occurrence. Then each $t_r\,(r \geqslant 1)$ has an exponential distribution with probability element $e^{-\lambda t}\lambda\,dt$.

Thus, starting with a binomial distribution of trials, we have reached an exponential distribution of times. It is to be expected, therefore, that there will be a close relation between (i) a Wald sequential test of the parameter p of a binomial population, considered in the limit when the values of p concerned are infinitesimally small, and (ii) a Wald sequential test of the parameter λ of an exponential population. It is the purpose of this note to point out that, given a test of either sort, (i) or (ii), there is a test of the other sort having an identical operating characteristic, and that the average sample sizes of these two tests are connected by a simple relation (given at (4) below).

For the binomial population, consider the likelihood-ratio sequential test (Wald, 1947) for the two hypotheses, H_0 that $p = p_0$ and H_1 that $p = p_1$, where $p_1 > p_0 > 0$. After any number n of observations have been taken, let n_1 denote the total number of occurrences of the event. Then sampling continues while

$$b + n \ln \frac{1-p_0}{1-p_1} < n_1 \ln \frac{p_1(1-p_0)}{p_0(1-p_1)} < a + n \ln \frac{1-p_0}{1-p_1}, \tag{1}$$

where $a\,(>0)$ and $b\,(<0)$ are constants, and sampling terminates as soon as one of the inequalities is violated, H_1 being accepted if the right-hand inequality is violated, H_0 if the left-hand inequality is violated. We call the chance of accepting H_0, considered as a function of the parameter p, the operating characteristic of the test. Note that if H_1 is accepted, the event occurred at the last observation, while if H_0 is accepted the event did not occur at the last observation.

Let us now set $p_0/\tau = \lambda_0$, $p_1/\tau = \lambda_1$, $n\tau = T$, and let $\tau \to 0$. The sampling condition (1) becomes

$$b + (\lambda_1 - \lambda_0)\,T < n_1 \ln (\lambda_1/\lambda_0) < a + (\lambda_1 - \lambda_0)\,T. \tag{1a}$$

Observation of the Poisson process continues as long as (1a) is satisfied, and ceases at the first instant when one of the inequalities in (1a) is violated. If the right-hand inequality is violated, so that H_1 is accepted, sampling terminates with an occurrence of the event, while if the left-hand inequality is violated and H_0 is accepted, sampling terminates with an interval in which the event does not occur. It is convenient to regard the operating characteristic as a function of λ, instead of p.

Turning now to an exponential population of times t with probability element $e^{-\lambda t}\lambda\,dt$, we consider the likelihood-ratio sequential test for the two hypotheses, H_0 that $\lambda = \lambda_0$ and H_1 that $\lambda = \lambda_1$, where $\lambda_1 > \lambda_0 > 0$. Sampling continues while

$$b' + (\lambda_1 - \lambda_0) \sum_{r=1}^{n} t_r < n \ln (\lambda_1/\lambda_0) < a' + (\lambda_1 - \lambda_0) \sum_{r=1}^{n} t_r, \tag{2}$$

where $a'\,(>0)$ and $b'\,(<0)$ are constants, and sampling terminates with the acceptance of H_1 if the right-hand inequality is first violated and with the acceptance of H_0 if the left-hand inequality is first violated. We call the chance of accepting H_0, as a function of λ, the operating characteristic of the test.

Let us now identify the observations t_r from the exponential population with the times between consecutive occurrences of the event in the Poisson process. The n in (2) is identified with the n_1 in (1a). The two tests, with respective conditions (1a) and (2), will lead always to the same decision, and so have the same operating characteristic, provided that

$$a' = a, \quad b' = b + \ln (\lambda_1/\lambda_0). \tag{3}$$

Then the difference between the tests is merely that if the observations lead to the acceptance of H_0, sampling rule (1a) requires that sampling ceases at the instant when the left-hand inequality becomes an equality, while sampling rule (2) prolongs the time of observation until the next occurrence of the event. If the observations lead to the acceptance of H_1, both rules operate alike. Note that the extent of observation T in rule (1a) corresponds to the cumulative sum of the readings Σt_r in rule (2), while the cumulative sum of the readings n_1 in rule (1a) corresponds to the number of observations n in rule (2).

Assuming always that (3) holds, let $\mathscr{E}_1$, $\mathscr{E}_{1a}$ and $\mathscr{E}_2$ denote expectations when sampling terminates according to rules (1), (1a) and (2), respectively. We have

$$\mathscr{E}_{1a}(T) = \lim_{\tau \to 0} \tau \mathscr{E}_1(n).$$

By a well-known identity of Wald's (1947, equation (A: 68)),

$$\mathscr{E}_{1a}(n_1) = \lambda \mathscr{E}_{1a}(T).$$

The n at the termination of sampling by rule (2) is equal either to n_1 or to $n_1 + 1$ at the termination of sampling by rule (1a), according as H_1 or H_0 is accepted, so that, if P denotes the probability that H_0 will be accepted,

$$\mathscr{E}_2(n) = \mathscr{E}_{1a}(n_1) + P.$$

Thus we have the relation

$$\mathscr{E}_2(n) = P + \lambda \mathscr{E}_{1a}(T) = P + \lim_{\tau \to 0} p \mathscr{E}_1(n). \tag{4}$$

Explicit formulae for P and for $\mathscr{E}_1(n)$ or $\mathscr{E}_{1a}(T)$ were given by Burman (1946). Thus explicit formulae are available for the operating characteristic and average sample size of the test of the exponential population. It may be noted that, apparently in ignorance of Burman's work, Dvoretzky, Kiefer & Wolfowitz (1953) have obtained the same formulae for the Poisson process.

The above results concerning the relation between the Poisson process and the exponential population can alternatively be deduced directly from a consideration of (i) the differential-difference equations satisfied by P and $\mathscr{E}_{1a}(T)$ for the Poisson process, these being limiting forms of the equations given by Burman, and (ii) the corresponding integral equations satisfied by P and $\mathscr{E}_2(n)$ for the exponential population.

REFERENCES

Burman, J. P. (1946). *Suppl. J. R. Statist. Soc.* **8**, 98–103.

Dvoretzky, A., Kiefer, J. & Wolfowitz, J. (1953). *Ann. Math. Statist.* **24**, 254–64.

Wald, A. (1947). *Sequential Analysis*. New York: Wiley.

Tables of generalized k-statistics

By S. H. ABDEL-ATY, *University College, London*

1. In a recent publication Wishart (1952) set out a technique for obtaining the moment coefficients of the multivariate distribution of the sample estimates (the k-statistics) of the population cumulants for the case where the population from which the sample is drawn is of finite size N. Use is made of a generalized k-statistic $k_{rs\ldots}$ having the property that its expectation over the finite population is equal to the same generalized K-statistic calculated from the whole population. This is expressed by

$$\mathscr{E}_N(k_{rs\ldots}) = K_{rs\ldots}. \tag{1·1}$$

$k_{rs\ldots}$ may be defined in terms of the augmented symmetric functions of the sample observations, and of the sample size n (see Wishart, 1952, equation (2·1)), or, inversely, any augmented symmetric function may be defined in terms of the $k_{rs\ldots}$, as in Wishart's equation (2·2). It follows that provided any sample criterion in which we are interested can be expressed in terms of monomial symmetric functions (in particular, this is true for polynomial functions of the Fisher k-statistics), its expectation can be obtained immediately by (1·1). Further, the generalized K-statistics in the answer may, if desired, be expressed in terms of the elements of the population by using the relation for $k_{rs\ldots}$ (Wishart, 1952, equation (2·1)), with K written for k and N for n, and interpreting the augmented symmetric functions as applying to the whole population.

2. Wishart provided a table giving the numerical coefficients of his reciprocal relations (2·1) and (2·2), up to and including functions of weight 6. The present table has been worked out for weight 12. To find a generalized k-statistic in terms of the augmented symmetric functions the entry is made *down* the appropriate column as far as and including the diagonal unity printed in bold type. For the inverse relationship the appropriate row is entered as far as and including the diagonal unity. Thus, for example, we find

$$k_{32^21^3} = 2[1^{12}]/n^{(12)} - 7[21^{10}]/n^{(11)} + 8[2^21^8]/n^{(10)} - 3[2^21^6]/n^{(9)} + [31^9]/n^{(10)} - 2[321^7]/n^{(9)} + [32^21^5]/n^{(8)}.$$

The table includes all tables of weight less than 12, in the sense that the required coefficients for a given weight can be picked out from it and the remainder discarded. Thus to obtain the table of weight 11 we cancel unity once in the subscript of the k's and in the augmented symmetrics and discard all columns and rows where it is not possible to do this. For example, from the expression for $k_{32^21^3}$ given above we obtain

$$k_{32^21^4} = 2[1^{11}]/n^{(11)} - 7[21^9]/n^{(10)} + 8[2^21^7]/n^{(9)} - 3[2^21^5]/n^{(8)} + [31^8]/n^{(9)} - 2[321^6]/n^{(8)} + [32^21^4]/n^{(7)}.$$

3. The introduction of the generalized k-statistics enables moments of sample criteria to be written down succinctly and it restores symmetry to the expressions for the case of the finite population where previously little or no pattern was visible. This is particularly so for the sample mean. Thus if we write

$$M(1^r) = \mathscr{E}_N(k_1 - K_1)^r,$$

$$A_r = \alpha^{r-1} - \frac{\alpha^{r-2}}{N} + \frac{\alpha^{r-3}}{N^2} - \ldots(-1)^{r-2}\frac{\alpha}{N^{r-2}},$$

$$\alpha = \frac{1}{n} - \frac{1}{N},$$

we have

$$M(1^1) = 0,$$

$$M(1^2) = K_2A_2,$$

$$M(1^3) = K_3A_3,$$

$$M(1^4) = K_4A_4 + 3K_{2^2}A_2^2,$$

$$M(1^5) = K_5A_5 + 10K_{32}A_3A_2,$$

$$M(1^6) = K_6A_6 + 15K_{42}A_4A_2 + 10K_{3^2}A_3^2 + 15K_{2^3}A_2^3,$$

$$M(1^7) = K_7A_7 + 21K_{52}A_5A_2 + 35K_{43}A_4A_3 + 105K_{32^2}A_3A_2^2,$$

$$M(1^8) = K_8A_8 + 28K_{62}A_6A_2 + 56K_{53}A_5A_3 + 35K_{4^2}A_4^2 + 210K_{42^2}A_4A_2^2 + 280K_{3^22}A_3^2A_2 + 105K_{2^4}A_2^4,$$

and so on. The numerical coefficients are the same as those of the appropriate augmented symmetric functions. Yet in spite of the elegance of these expressions it is open to question whether the procedure involved in applying them to actual data is any improvement over that suggested by Irwin & Kendall (1944). In order to obtain $M(1^8)$ in terms of the moments of the population it is necessary to express the generalized K-statistics in terms of Fisher's K's. Wishart gives the necessary relations, but the solution of a determinant of the sixth order is required, the same determinant incidentally as that to which we are led by Irwin & Kendall. This would appear always to be the case.

I should like to express my thanks to Dr F. N. David and Professor M. G. Kendall for the encouragement they have given me in this work.

REFERENCES

Irwin, J. O. & Kendall, M. G. (1944). *Ann. Eugen., Lond.*, **12**, 138.
Wishart, J. (1952). *Biometrika*, **39**, 1.

Weight 12 (i)	h_{12}	h_{10}	h_{21^2}	h_{2^3}	h_{31^3}	$h_{2^21^2}$	h_{2^4}	h_{41^4}	h_{321}	$h_{2^31^2}$	h_{421}	h_{31^5}	$h_{2^21^4}$	h_{21^6}
$[1^{12}]/n^{(12)}$	1	−1	1	−1	1	−1	1	2	−2	2	−2	2	4	−4	4
$[21^{10}]/n^{(11)}$	1	1	−2	3	−4	5	−6	−3	5	−7	9	−11	−12	16	−20
$[2^21^8]/n^{(10)}$	1	2	1	−3	6	−10	15	.	−3	8	−15	24	9	−21	37
$[2^31^6]/n^{(9)}$	1	3	3	1	−4	10	−20	.	.	−3	11	−26	.	9	−30
$[2^41^4]/n^{(8)}$	1	4	6	4	1	−5	15	.	.	.	−3	14	.	.	9
$[2^51^2]/n^{(7)}$	1	5	10	10	5	1	−6	.	.	.	.	−3	.	.	.
$[2^6]/n^{(6)}$	1	6	15	20	15	6	1	.	.	.	.	.	.	.	.
$[31^9]/n^{(10)}$	1	3	.	.	.	.	.	1	−1	1	−1	1	4	−4	4
$[321^7]/n^{(9)}$	1	4	3	.	.	.	.	1	1	−2	3	−4	−6	10	−14
$[32^21^5]/n^{(8)}$	1	5	7	3	.	.	.	1	2	1	−3	6	.	−6	16
$[32^31^3]/n^{(7)}$	1	6	12	10	3	.	.	1	3	3	1	−4	.	.	−6
$[32^41]/n^{(6)}$	1	7	18	22	13	3	.	1	4	6	4	1	.	.	.
$[3^21^6]/n^{(6)}$	1	6	9	.	.	.	.	2	4	.	.	.	1	−1	1
$[3^221^4]/n^{(7)}$	1	7	15	9	.	.	.	2	5	4	1	.	.	1	−2
$[3^231^2]/n^{(8)}$	1	8	22	24	9	.	.	2	8	6	.	.	1	1	1
$[3^241]/n^{(9)}$	1	9	30	40	33	9	.	2	10	14	6	.	1	2	1
$[3^251]/n^{(8)}$	1	9	27	27	.	.	.	2	12	24	20	6	1	3	3
$[3^261]/n^{(8)}$	1	10	36	54	27	.	.	3	18	27	.	.	3	9	.
$[3^6]/n^{(6)}$	1	12	54	108	81	.	.	3	21	45	27	.	3	12	9
$[41^8]/n^{(9)}$	1	6	3	.	.	.	.	4	6	.	.	.	6	36	54
$[421^5]/n^{(8)}$	1	7	9	3	.	.	.	4	4	.	.	.	.	.	.
$[42^21^4]/n^{(7)}$	1	8	10	12	.	.	.	4	8	4	.	.	.	.	.
$[42^31^2]/n^{(6)}$	1	9	24	28	3	.	.	4	12	12	4	.	.	.	.
$[42^41]/n^{(6)}$	1	10	33	52	43	18	3	4	16	24	16	4	.	.	.
$[431^7]/n^{(7)}$	1	9	21	9	.	.	.	5	18	3	.	.	4	.	.
$[4321^5]/n^{(6)}$	1	10	30	30	9	.	.	5	23	21	3	.	4	4	.
$[432^21^3]/n^{(5)}$	1	11	40	60	39	9	.	5	28	44	24	3	4	8	4
$[43^21^4]/n^{(5)}$	1	12	48	72	27	.	.	6	42	78	18	.	9	30	3
$[43^241]/n^{(6)}$	1	13	60	120	99	27	.	6	48	120	96	18	9	39	33
$[4^21^6]/n^{(6)}$	1	12	42	36	9	.	.	8	48	24	.	.	16	.	.
$[4^221^4]/n^{(5)}$	1	13	54	78	45	9	.	8	56	72	24	.	16	16	.
$[4^231^2]/n^{(4)}$	1	14	67	132	123	54	9	8	64	128	96	24	16	32	16
$[4^241]/n^{(4)}$	1	15	78	162	117	27	.	9	84	210	108	9	24	96	24
$[4^251]/n^{(5)}$	1	18	117	324	351	162	27	12	144	504	432	108	48	288	144
$[521^6]/n^{(6)}$	1	10	15	.	.	.	.	10	10	.	.	.	.	.	.
$[52^21^5]/n^{(7)}$	1	11	25	15	.	.	.	10	20	10	.	.	.	.	.
$[52^31^3]/n^{(6)}$	1	12	36	40	15	.	.	10	30	30	10	.	.	.	.
$[52^41]/n^{(6)}$	1	13	48	76	55	15	.	10	40	60	40	10	.	.	.
$[531^5]/n^{(6)}$	1	13	45	45	.	.	.	11	50	45	.	.	10	10	.
$[5321^4]/n^{(6)}$	1	14	58	90	45	.	.	11	61	95	45	.	10	20	10
$[532^21^2]/n^{(4)}$	1	15	72	148	135	45	.	11	72	156	140	45	10	30	30
$[53^21]/n^{(4)}$	1	16	84	180	135	.	.	12	96	240	180	.	21	90	75
$[541^4]/n^{(5)}$	1	16	78	120	45	.	.	14	110	150	30	.	40	40	.
$[5421]/n^{(4)}$	1	17	94	198	165	45	.	14	124	260	180	30	40	80	40
$[543]/n^{(3)}$	1	19	126	354	405	135	.	15	168	558	600	135	54	270	270
$[54^21^2]/n^{(4)}$	1	20	130	300	225	.	.	20	220	500	300	.	100	200	100
$[6^21^4]/n^{(4)}$	1	21	150	430	525	225	.	20	240	720	800	300	100	300	300
$[61^6]/n^{(7)}$	1	15	45	15	.	.	.	20	60	.	.	.	10	.	.
$[621^4]/n^{(6)}$	1	16	60	60	15	.	.	20	80	60	.	.	10	10	.
$[62^21^2]/n^{(5)}$	1	17	76	120	75	15	.	20	100	140	60	.	10	20	10
$[62^3]/n^{(4)}$	1	18	93	196	195	90	15	20	120	240	200	60	10	30	30
$[631^3]/n^{(5)}$	1	18	90	150	45	.	.	21	135	225	15	.	30	90	.
$[6321]/n^{(4)}$	1	19	108	240	195	45	.	21	156	360	240	15	30	120	90
$[63^21]/n^{(3)}$	1	21	144	420	495	135	.	22	216	720	840	90	51	315	495
$[641^2]/n^{(4)}$	1	21	138	330	225	45	.	24	240	600	240	.	90	300	30
$[642]/n^{(3)}$	1	22	159	468	555	270	45	24	264	840	840	240	90	390	330
$[651]/n^{(3)}$	1	25	210	690	825	225	.	30	420	1500	1500	150	210	900	750
$[6^3]/n^{(3)}$	1	30	315	1380	2475	1350	225	40	720	3600	6000	1800	420	2700	4500
$[71^5]/n^{(6)}$	1	21	105	105	.	.	.	35	210	105	.	.	70	.	.
$[721^3]/n^{(5)}$	1	22	126	210	105	.	.	35	245	315	105	.	70	70	.
$[72^21]/n^{(4)}$	1	23	148	336	315	105	.	35	280	560	420	105	70	140	70
$[731^2]/n^{(4)}$	1	24	168	420	315	.	.	36	336	840	420	.	105	420	105
$[732]/n^{(3)}$	1	25	192	588	735	315	.	36	372	1176	1260	420	105	525	525
$[741]/n^{(3)}$	1	27	234	798	945	315	.	39	504	1890	1680	315	210	1260	630
$[75]/n^{(2)}$	1	31	330	1470	2625	1575	.	45	780	3990	6300	2625	420	3150	4200
$[81^4]/n^{(5)}$	1	28	210	420	105	.	.	56	560	840	.	.	280	280	.
$[821^2]/n^{(4)}$	1	29	238	630	525	105	.	56	616	1400	840	.	280	560	280
$[82^2]/n^{(3)}$	1	30	267	868	1155	630	105	56	672	2016	2240	840	280	840	840
$[831]/n^{(3)}$	1	31	294	1050	1365	315	.	57	756	2730	2940	105	336	1680	1680
$[84]/n^{(2)}$	1	34	381	1764	3255	1890	315	60	1008	5208	8400	2940	504	4200	5880
$[91^3]/n^{(4)}$	1	36	378	1260	945	.	.	84	1260	3780	1260	.	840	2520	.
$[921]/n^{(3)}$	1	37	414	1638	2205	945	.	84	1344	5040	5040	1260	840	3360	2520
$[93]/n^{(2)}$	1	39	480	2394	4725	2835	.	85	1548	7938	13860	4725	924	6300	11340
$[10, 1^2]/n^{(2)}$	1	45	630	3150	4725	945	.	120	2520	12600	12600	.	2100	12600	6300
$[10, 2]/n^{(2)}$	1	46	675	3780	7875	5670	945	120	2640	15120	25200	12600	2100	14700	18900
$[11, 1]/n^{(2)}$	1	55	990	6930	17325	10395	.	165	4620	34650	69300	17325	4620	46200	69300
$[12]/n$	1	66	1485	13860	51975	62370	10395	220	7920	83160	277200	207900	9240	138600	415800

Weight 12 (ii)	$k_{0^6 2^6}$	$k_{0^6 2^4 1^2}$	$k_{0^6 2^2 1^4}$	$k_{0^6 1^6}$	$k_{1^2 2^5}$	$k_{1^2 2^3 1^3}$	$k_{1^2 2 1^5}$	$k_{1^4 2^4}$	$k_{1^4 2^2 1^4}$	$k_{1^4 1^8}$	$k_{1^6 2^3}$	$k_{1^6 2 1^3}$	$k_{1^8 2^2}$	$k_{1^8 2 1^2}$	$k_{1^{10} 2}$
$[1^{12}]/n^{(12)}$	−4	8	−8	16	−6	6	−6	6	−6	−12	12	−12	−24		
$21^{10}]/n^{(11)}$	24	−36	44	−96	12	−18	24	−30	36	42	−54	66	120		
$2^2 1^8]/n^{(10)}$	−57	54	−90	216	−3	15	−33	57	−87	−42	84	−138	−210		
$2^3 1^6]/n^{(9)}$	67	−27	81	−216	.	−3	18	−51	108	9	−51	135	144		
$2^4 1^4]/n^{(8)}$	−39	.	−27	81	.	.	−3	21	−72	.	9	−60	−27		
$2^5 1^2]/n^{(7)}$	9	.	.	.	.	.	.	−3	24	.	.	9	.		
$2^6]/n^{(6)}$	.	.	.	.	.	.	.	.	−3	.	.	.	.		
$31^9]/n^{(10)}$	−4	12	−12	32	−4	4	−4	4	−4	−14	14	−14	−40		
$321^7]/n^{(9)}$	18	−36	48	−144	.	−4	8	−12	16	24	−38	52	132		
$32^2 1^3]/n^{(8)}$	−30	27	−63	216	.	.	−4	12	−24	−3	27	−65	−120		
$[32^2 1^5]/n^{(7)}$	22	.	27	−108	.	.	.	−4	16	.	−3	30	18		
$32^4 1]/n^{(6)}$	−6	.	.	.	.	.	.	.	−4	.	.	−3	.		
$3^2 1^6]/n^{(6)}$	−1	6	−6	24	.	.	.	.	.	−4	4	−4	−22		
$3^2 21^4]/n^{(7)}$	3	−9	15	−72	.	.	.	.	.	.	−4	8	36		
$3^2 2^2 1^2]/n^{(6)}$	−3	.	−9	54	.	.	.	.	.	.	.	−4	−3		
$3^2 2^4]/n^{(6)}$	1	.	.	.	.	.	.	.	.	.	.	.	.		
$3^3 1^3]/n^{(6)}$	.	1	−1	8	.	.	.	.	.	.	.	.	−4		
$3^3 21]/n^{(6)}$	.	1	1	−12	.	.	.	.	.	.	.	.	.		
$3^4]/n^{(6)}$	.	4	12	1	.	.	.	.	.	.	.	.	.		
$41^5]/n^{(6)}$	.	.	.	.	1	−1	1	−1	1	2	−2	2	4		
$[421^4]/n^{(6)}$	.	.	.	.	1	1	−2	3	−4	−3	5	−7	−12		
$42^2 1^3]/n^{(7)}$	.	.	.	.	1	2	1	−3	6	.	−3	8	9		
$42^3 1^2]/n^{(6)}$	.	.	.	.	1	3	3	1	−4	.	.	−3	.		
$42^4]/n^{(6)}$	.	.	.	.	1	4	6	4	1	.	.	.	.		
$431^3]/n^{(7)}$	.	.	.	.	1	3	.	.	.	1	−1	1	4		
$4321^2]/n^{(6)}$	.	.	.	.	1	4	3	.	.	1	1	−2	−6		
$432^2 1]/n^{(6)}$	.	.	.	.	1	5	7	3	.	1	2	1	.		
$43^2 1^2]/n^{(6)}$	.	4	.	.	1	6	9	.	.	2	6	.	1		
$43^2 2]/n^{(6)}$	3	4	4	.	1	7	15	9	.	2	8	6	1		
$4^2 1^4]/n^{(6)}$	.	.	.	.	2	12	6	.	.	8	.	.	.		
$[4^2 2^2]/n^{(6)}$	.	.	.	.	2	14	18	6	.	8	8	.	.		
$4^3 2]/n^{(6)}$	.	.	.	.	2	16	32	24	6	8	16	8	.		
$4^3 21]/n^{(6)}$	.	16	.	.	2	18	42	18	.	10	36	6	8		
$4^4]/n^{(6)}$	.	64	.	.	3	36	126	108	27	24	144	72	48		
$51^7]/n^{(8)}$	.	.	.	.	5	.	.	.	.	.	.	.	.		
$521^5]/n^{(7)}$	.	.	.	.	5	5	.	.	.	.	.	.	.		
$52^2 1^3]/n^{(6)}$	.	.	.	.	5	10	5	.	.	.	.	.	.		
$52^3 1]/n^{(6)}$	.	.	.	.	5	15	15	5	.	.	.	.	.		
$531^4]/n^{(6)}$	.	.	.	.	5	15	.	.	.	5	.	.	.		
$5321^3]/n^{(6)}$	.	.	.	.	5	20	15	.	.	5	5	.	.		
$[532^2]/n^{(6)}$	10	.	.	.	5	25	35	15	.	5	10	5	.		
$53^2 1]/n^{(6)}$	.	10	10	.	5	30	45	.	.	10	30	.	5		
$541^3]/n^{(6)}$	.	.	.	.	6	40	30	.	.	30	10	.	.		
$5421]/n^{(6)}$	.	.	.	.	6	46	70	30	.	30	40	10	.		
$543]/n^{(6)}$	30	40	40	.	6	58	150	90	.	36	140	60	30		
$5^2 1^2]/n^{(6)}$	.	.	.	.	10	100	150	.	.	100	100	.	.		
$5^2 2]/n^{(6)}$	100	.	.	.	10	110	250	150	.	100	200	100	.		
$61^6]/n^{(7)}$	.	.	.	.	15	15	.	.	.	.	.	.	.		
$621^4]/n^{(6)}$	.	.	.	.	15	30	15	.	.	.	.	.	.		
$62^2 1^2]/n^{(6)}$	.	.	.	.	15	45	45	15	.	.	.	.	.		
$[62^3]/n^{(6)}$	10	.	.	.	15	60	90	60	15	.	.	.	.		
$631^3]/n^{(6)}$	.	10	.	.	15	60	45	.	.	15	15	.	.		
$6321]/n^{(6)}$	.	10	10	.	15	75	105	45	.	15	30	15	.		
$63^2]/n^{(6)}$	15	40	120	10	15	105	225	135	.	30	120	90	15		
$641^2]/n^{(6)}$	.	40	.	.	16	120	180	60	.	80	120	.	10		
$642]/n^{(6)}$	30	40	40	.	16	136	300	240	60	80	200	120	10		
$651]/n^{(6)}$	.	100	100	.	20	240	600	300	.	250	600	150	50		
$6^2]/n^{(6)}$	300	400	1200	100	30	480	1800	1800	450	600	2400	1800	300		
$71^5]/n^{(6)}$	.	.	.	.	35	105	.	.	.	35	.	.	.		
$721^3]/n^{(6)}$	.	.	.	.	35	140	105	.	.	35	35	.	.		
$[72^2 1]/n^{(6)}$	.	.	.	.	35	175	245	105	.	35	70	35	.		
$731^2]/n^{(6)}$	.	70	.	.	35	210	315	.	.	70	210	.	35		
$732]/n^{(6)}$	105	70	70	.	35	245	525	315	.	70	280	210	35		
$741]/n^{(6)}$	.	280	.	.	36	336	840	420	.	210	840	210	210		
$75]/n^{(6)}$	1050	700	700	.	40	500	2100	2100	.	560	2800	2100	700		
$81^4]/n^{(6)}$	.	.	.	.	70	420	210	.	.	280	.	.	.		
$821^2]/n^{(6)}$	.	.	.	.	70	490	630	210	.	280	280	.	.		
$82^2]/n^{(6)}$	280	.	.	.	70	560	1120	840	210	280	560	280	.		
$831]/n^{(6)}$	.	280	280	.	70	630	1470	630	.	350	1260	210	280		
$84]/n^{(6)}$	840	1120	1120	.	71	868	3150	2940	735	616	3920	2520	1400		
$[91^3]/n^{(6)}$	.	280	.	.	126	1260	1890	.	.	1260	1260	.	.		
$921]/n^{(6)}$	.	280	280	.	126	1386	3150	1890	.	1260	2520	1260	.		
$93]/n^{(6)}$	1260	1120	3360	280	126	1638	5670	5670	.	1386	6300	5670	1260		
$[10, 1^2]/n^{(6)}$	.	2800	.	.	210	3150	9450	3150	.	4200	12600	.	2100		
$[10, 2]/n^{(6)}$	6300	2800	2800	.	210	3360	12600	12600	3150	4200	16800	12600	2100		
$[11, 1]/n^{(6)}$	.	15400	15400	.	330	6930	34650	34650	.	11550	69300	34650	23100		
$[12]/n$	138600	61600	184800	15400	495	13860	103950	207900	51975	27720	277200	415800	138600		

Weight 12 (iii)	$k_{10^2 2}$	$k_{9^3 1^4}$	$k_{8^3 2^2 1^2}$	$k_{8^2 3^2}$	$k_{8^2 2^4}$	$k_{7^3 5}$	$k_{6^2 1^7}$	$k_{6^2 2^3}$	$k_{6^2 3^2 1^2}$	$k_{6^2 4^2}$	$k_{6^2 2^2 1^4}$	$k_{6^2 2^6}$	$k_{6^3 2^3}$
$[1^{12}]/n^{(12)}$	24	36	−36	36	72	−216	24	−24	24	−24	48	−48	48
$[21^{10}]/n^{(11)}$	−144	−144	180	−216	−396	1296	−60	84	−108	132	−192	240	−288
$[2^2 1^8]/n^{(10)}$	330	180	−324	504	792	−2916	30	−90	174	−282	240	−432	672
$[2^3 1^6]/n^{(9)}$	−354	−72	252	−576	−684	3024	.	30	−120	294	−90	330	−762
$[2^4 1^4]/n^{(8)}$	171	9	−81	333	234	−1458	.	.	30	−150	.	−90	420
$[2^5 1^2]/n^{(7)}$	−27	.	9	−90	−27	324	.	.	.	30	.	.	−90
$[2^6]/n^{(6)}$	.	.	.	9	.	−27	.	.	.	.	.	.	.
$[31^9]/n^{(10)}$	40	48	−48	48	132	−432	20	−20	20	−20	64	−64	64
$[321^7]/n^{(9)}$	−172	−96	144	−192	−480	1728	−10	30	−50	70	−140	204	−268
$[32^2 1^5]/n^{(8)}$	252	24	−120	264	516	−2160	.	−10	40	−90	60	−200	404
$[32^3 1^3]/n^{(7)}$	−138	.	24	−144	−144	864	.	.	−10	50	.	60	−260
$[32^4 1]/n^{(6)}$	18	.	.	24	9	−108	.	.	.	−10	.	.	60
$[3^2 1^6]/n^{(8)}$	22	16	−16	16	80	−288	.	.	.	.	20	−20	20
$[3^2 21^4]/n^{(7)}$	−58	.	16	−32	−144	576	.	.	.	.	−10	30	−50
$[3^2 2^2 1^2]/n^{(6)}$	39	.	.	16	24	−144	.	.	.	.	.	−10	40
$[3^2 2^3]/n^{(5)}$	−3	.	.	.	.	.	.	.	.	.	.	.	−10
$[3^3 1^3]/n^{(6)}$	4	.	.	.	16	−64	.	.	.	.	.	.	.
$[3^3 21]/n^{(5)}$	−4	.	.	.	.	.	.	.	.	.	.	.	.
$[3^4]/n^{(4)}$	.	.	.	.	.	.	.	.	.	.	.	.	.
$[41^8]/n^{(9)}$	−4	−12	12	−12	−24	108	−3	5	−5	5	−10	10	−10
$[421^6]/n^{(8)}$	16	24	−36	48	84	−432	.	−5	10	−15	15	−25	35
$[42^2 1^4]/n^{(7)}$	−21	−6	30	−66	−84	540	.	.	−5	15	.	15	−40
$[42^3 1^2]/n^{(6)}$	9	.	−6	36	18	−216	.	.	.	−5	.	.	15
$[42^4]/n^{(5)}$	.	.	.	−6	.	27	.	.	.	.	.	.	.
$[431^5]/n^{(7)}$	−4	−8	.	−8	−28	144	.	.	.	.	−5	5	−5
$[4321^3]/n^{(6)}$	10	.	−8	16	48	−288	.	.	.	.	.	−5	10
$[432^2 1]/n^{(5)}$	−6	.	.	−8	−6	72	.	.	.	.	.	.	−5
$[43^2 1^3]/n^{(5)}$	−1	.	.	.	−8	48	.	.	.	.	.	.	.
$[43^2 2]/n^{(4)}$	1	.	.	.	.	.	.	.	.	.	.	.	.
$[4^2 1^4]/n^{(6)}$	.	1	−1	1	2	−18	.	.	.	.	.	.	.
$[4^2 21^2]/n^{(6)}$	.	1	1	−2	−3	36	.	.	.	.	.	.	.
$[4^2 2^2]/n^{(6)}$	.	1	2	1	.	−9	.	.	.	.	.	.	.
$[4^3 31]/n^{(6)}$	.	1	3	.	1	−12	.	.	.	.	.	.	.
$[4^3]/n^{(5)}$	.	3	18	9	12	1	.	.	.	.	.	.	.
$[51^7]/n^{(9)}$	.	.	.	.	.	.	1	−1	1	−1	2	−2	2
$[521^5]/n^{(7)}$	.	.	.	.	.	.	1	1	−2	3	−3	5	−7
$[52^2 1^3]/n^{(6)}$	.	.	.	.	.	.	1	2	1	−3	.	−3	8
$[52^3 1]/n^{(5)}$	.	.	.	.	.	.	1	3	3	1	.	.	−3
$[531^4]/n^{(6)}$	.	.	.	.	.	.	1	3	.	.	1	−1	1
$[5321^2]/n^{(6)}$	.	.	.	.	.	.	1	4	3	.	1	1	−2
$[532^2]/n^{(6)}$	.	.	.	.	.	.	1	5	7	3	1	2	1
$[53^2 1]/n^{(6)}$	.	.	.	.	.	.	1	6	9	.	2	6	.
$[541^3]/n^{(6)}$	.	5	.	.	.	.	1	6	3	.	4	.	.
$[542]/n^{(4)}$	.	5	5	.	.	.	1	7	9	3	4	4	.
$[543]/n^{(3)}$	10	5	15	.	5	.	1	9	21	9	5	18	3
$[5^2 1^4]/n^{(6)}$	.	25	.	.	.	.	2	20	30	.	20	20	.
$[5^2 2]/n^{(3)}$	.	25	25	.	.	.	2	22	50	30	20	40	20
$[61^6]/n^{(7)}$	.	.	.	.	.	.	6	.	.	.	.	.	.
$[621^4]/n^{(6)}$	.	.	.	.	.	.	6	6	.	.	.	.	.
$[62^2 1^2]/n^{(5)}$	.	.	.	.	.	.	6	12	6	.	.	.	.
$[62^3]/n^{(4)}$	.	.	.	.	.	.	6	18	18	6	.	.	.
$[631^3]/n^{(5)}$	.	.	.	.	.	.	6	18	.	.	6	.	.
$[6321]/n^{(4)}$	.	.	.	.	.	.	6	24	18	.	6	6	.
$[63^2]/n^{(3)}$	15	.	.	.	.	.	6	36	54	.	12	36	.
$[641^2]/n^{(4)}$	.	15	15	.	.	.	6	36	18	.	24	.	.
$[642]/n^{(3)}$	10	15	30	15	.	.	6	42	54	18	24	24	.
$[651]/n^{(3)}$	.	75	75	.	.	.	7	75	135	15	80	120	.
$[6^2]/n^{(2)}$	300	225	450	225	.	.	12	180	540	180	240	720	.
$[71^5]/n^{(6)}$	.	.	.	.	.	.	21	21	.	.	.	.	.
$[721^3]/n^{(5)}$	.	.	.	.	.	.	21	42	21	.	.	.	.
$[72^2 1]/n^{(4)}$	.	.	.	.	.	.	21	63	63	21	.	.	.
$[731^2]/n^{(4)}$	.	.	.	.	.	.	21	84	63	.	21	21	.
$[732]/n^{(3)}$	35	.	.	.	.	.	21	105	147	63	21	42	21
$[741]/n^{(3)}$	.	35	105	.	35	.	21	147	189	63	84	84	.
$[75]/n^{(2)}$	350	175	525	.	175	.	22	252	630	420	245	630	315
$[81^4]/n^{(5)}$	.	35	.	.	.	.	56	168	.	.	56	.	.
$[821^2]/n^{(4)}$	.	35	35	.	.	.	56	224	168	.	56	56	.
$[82^2]/n^{(3)}$	.	35	70	35	.	.	56	280	392	168	56	112	56
$[831]/n^{(3)}$	.	35	105	.	35	.	56	336	504	.	112	336	.
$[84]/n^{(2)}$	280	105	630	315	420	35	56	504	1176	504	280	1008	168
$[91^3]/n^{(4)}$	.	315	.	.	.	.	126	756	378	.	504	.	.
$[921]/n^{(3)}$	.	315	315	.	.	.	126	882	1134	378	504	504	.
$[93]/n^{(2)}$	1260	315	945	.	315	.	126	1134	2646	1134	630	2268	378
$[10, 1^2]/n^{(3)}$	.	1575	1575	.	.	.	252	2520	3780	.	2520	2520	.
$[10, 2]/n^{(2)}$	2100	1575	3150	1575	.	.	252	2772	6300	3780	2520	5040	2520
$[11, 1]/n^{(2)}$	.	5775	17325	.	5775	.	462	6930	20790	6930	9240	27720	.
$[12]/n$	138600	17325	103950	51975	69300	5775	792	16632	83160	83160	27720	166320	83160

Weight 12 (iv)	$h_{12}b_1$	$h_{111}{}^2$	h_{1111}	h_{222}	$h_6b_1{}^2$	h_6b_2	$h_{21}{}^2$	$h_{2211}{}^2$	$h_{21}b_1{}^2$	$h_{222}{}^1$	$h_{221}{}^2$	h_{2211}	$h_{33}{}^1$
$[1^{12}]/n^{(12)}$	96	−144	144	−288	576	−576	−120	120	−120	120	−240	240	−480
$21^{10}]/n^{(11)}$	−528	648	−792	1728	−2880	3456	360	−480	600	−720	1080	−1320	2880
$2^21^8]/n^{(10)}$	1056	−972	1620	−3888	5040	−7920	−270	630	−1110	1710	−1620	2700	−6480
$2^21^6]/n^{(9)}$	−900	540	−1512	3996	−3600	8640	30	−300	930	−2040	870	−2490	6600
$2^21^4]/n^{(8)}$	270	−90	630	−1800	900	−4500	.	30	−330	1260	−90	900	−2790
$2^21^2]/n^{(7)}$	.	.	−90	270	.	900	.	.	30	−360	.	−90	270
$2^6]/n^{(6)}$	.	.	.	.	.	.	.	.	.	30	.	.	.
$31^9]/n^{(10)}$	176	−216	216	−576	960	−960	−120	120	−120	120	−360	360	−960
$321^7]/n^{(9)}$	−664	540	−756	2376	−2880	3840	120	−240	360	−480	960	−1320	4080
$32^21^3]/n^{(8)}$	780	−300	840	−3192	2400	−5280	.	120	−360	720	−630	1590	−5760
$[32^21^5]/n^{(7)}$	−270	30	−330	1500	−600	3000	.	.	120	−480	30	−660	2820
$32^41]/n^{(6)}$	.	.	30	−180	.	−600	.	.	.	120	.	30	−180
$3^21^6]/n^{(8)}$	104	−80	80	−376	400	−400	−10	10	−10	10	−140	140	−640
$3^221^4]/n^{(7)}$	−220	40	−120	860	−400	800	.	−10	20	−30	150	−290	1680
$3^22^21^2]/n^{(6)}$	90	.	40	−420	100	−500	.	.	−10	30	.	150	−1080
$3^22^6]/n^{(6)}$	.	.	.	30	.	100	.	.	.	−10	.	.	30
$3^31^3]/n^{(6)}$	20	.	.	−80	.	.	.	.	.	.	−10	10	−160
$3^321]/n^{(6)}$	−10	.	.	40	.	.	.	.	.	.	.	−10	180
$3^4]/n^{(6)}$	.	.	.	.	.	.	.	.	.	.	.	.	−10
$41^8]/n^{(9)}$	−20	54	−54	108	−240	240	30	−30	30	−30	60	−60	120
$[421^6]/n^{(8)}$	60	−120	174	−402	600	−840	−15	45	−75	105	−120	180	−420
$42^21^4]/n^{(7)}$	−45	45	−165	450	−300	900	.	−15	60	−135	45	−165	450
$42^31^2]/n^{(6)}$	.	.	45	−135	.	−300	.	.	−15	75	.	45	−135
$42^4]/n^{(6)}$	.	.	.	.	.	.	.	.	.	−15	.	.	.
$431^5]/n^{(7)}$	−20	40	−40	134	−200	200	.	.	.	.	30	−30	120
$4321^3]/n^{(6)}$	30	−10	50	−260	100	−300	.	.	.	.	−15	45	−240
$432^21]/n^{(5)}$	.	.	−10	75	.	100	.	.	.	.	.	−15	90
$43^31^2]/n^{(6)}$	−5	.	.	−40	.	.	.	.	.	.	.	.	30
$43^32]/n^{(4)}$	.	.	.	−10	.	.	.	.	.	.	.	.	−15
$4^21^4]/n^{(6)}$	.	−5	5	−10	25	−25	.	.	.	.	.	.	.
$4^221^2]/n^{(5)}$	.	.	−5	15	.	25	.	.	.	.	.	.	.
$4^22^2]/n^{(4)}$	.	.	.	.	.	.	.	.	.	.	.	.	.
$4^31]/n^{(4)}$	.	.	.	−5	.	.	.	.	.	.	.	.	.
$4^4]/n^{(4)}$	.	.	.	.	.	.	.	.	.	.	.	.	.
$51^7]/n^{(8)}$	4	−6	6	−12	48	−48	−6	6	−6	6	−12	12	−24
$521^5]/n^{(7)}$	−12	12	−18	42	−120	108	.	−6	12	−18	18	−30	72
$52^21^3]/n^{(6)}$	9	−3	15	−42	60	−180	.	.	−6	18	.	18	−54
$52^31]/n^{(5)}$	.	.	−3	9	.	60	.	.	.	−6	.	.	.
$531^4]/n^{(6)}$	4	−4	4	−14	40	−40	.	.	.	.	−6	6	−24
$5321^2]/n^{(5)}$	−6	.	−4	24	−20	60	.	.	.	.	.	−6	36
$[532^2]/n^{(4)}$	.	.	.	−3	.	−20	.	.	.	.	.	.	.
$53^21]/n^{(4)}$	x	.	.	−4	.	.	.	.	.	.	.	.	−6
$541^3]/n^{(5)}$	.	x	−1	2	−10	10	.	.	.	.	.	.	.
$5421]/n^{(4)}$	.	1	x	−3	.	−10	.	.	.	.	.	.	.
$543]/n^{(3)}$	4	1	3	x	.	.	.	.	.	.	.	.	.
$5^21^3]/n^{(4)}$	.	10	.	.	x	−1	.	.	.	.	.	.	.
$5^32]/n^{(3)}$	.	10	10	.	1	x	.	.	.	.	.	.	.
$61^6]/n^{(7)}$	.	.	.	.	.	.	x	−1	1	−1	2	−2	4
$621^4]/n^{(6)}$	.	.	.	.	.	.	1	x	−2	3	−3	5	−12
$62^21^2]/n^{(5)}$	.	.	.	.	.	.	1	2	x	−3	.	−3	9
$[62^3]/n^{(4)}$	.	.	.	.	.	.	1	3	3	x	.	.	.
$631^3]/n^{(5)}$	.	.	.	.	.	.	1	3	.	.	x	−1	4
$6321]/n^{(4)}$	.	.	.	.	.	.	1	4	3	.	1	x	−6
$63^2]/n^{(3)}$	6	.	.	.	.	.	1	6	9	.	2	6	x
$641^2]/n^{(4)}$	.	6	.	.	.	.	1	6	3	.	4	.	.
$643]/n^{(3)}$	.	6	6	.	.	.	1	7	9	3	4	4	.
$651]/n^{(3)}$	10	45	15	.	6	.	1	10	15	.	10	10	.
$6^2]/n^{(3)}$	120	180	180	.	36	.	2	30	90	30	40	120	20
$71^5]/n^{(6)}$	.	.	.	.	.	.	7	.	.	.	.	.	.
$721^3]/n^{(5)}$	.	.	.	.	.	.	7	7	.	.	.	.	.
$[72^21]/n^{(4)}$	.	.	.	.	.	.	7	14	7	.	.	.	.
$731^2]/n^{(4)}$	.	.	.	.	.	.	7	21	.	.	7	.	.
$732]/n^{(3)}$	.	.	.	.	.	.	7	28	21	.	7	7	.
$741]/n^{(3)}$	.	21	21	.	.	.	7	42	21	.	28	.	.
$75]/n^{(2)}$	70	140	210	35	21	21	7	70	105	.	70	70	.
$81^4]/n^{(5)}$	.	.	.	.	.	.	28	28	.	.	.	.	.
$821^2]/n^{(4)}$	.	.	.	.	.	.	28	56	28	.	.	.	.
$82^2]/n^{(3)}$	.	.	.	.	.	.	28	84	84	28	.	.	.
$831]/n^{(3)}$	56	.	.	.	.	.	28	112	84	.	28	28	.
$84]/n^{(2)}$	224	56	168	56	.	.	28	196	252	84	112	112	.
$[91^7]/n^{(6)}$	.	126	.	.	.	.	84	252	.	.	84	.	.
$921]/n^{(3)}$	.	126	126	.	.	.	84	336	252	.	84	84	.
$93]/n^{(2)}$	504	126	378	126	.	.	84	504	756	.	168	504	84
$10, 1^2]/n^{(3)}$	.	1260	.	.	126	.	210	1260	630	.	840	.	.
$10, 2]/n^{(2)}$	.	1260	1260	.	126	126	210	1470	1890	630	840	840	.
$11, 1]/n^{(2)}$	4620	6930	6930	.	1386	.	462	4620	6930	.	4620	4620	.
$12]/n$	55440	27720	83160	27720	8316	8316	924	13860	41580	13860	18480	55440	9240

Weight 12 (v)	k_{6211}	k_{6111}	k_{621}	k_{42}	k_{72}	k_{7211}	k_{7112}	k_{7111}	k_{7101}	k_{711}	k_{72}	k_{621}	k_{6111}
$[1^{12}]/n^{(12)}$	720	−720	−2880	14400	720	−720	720	1440	−1440	−4320	17280	−5040	5040
$21^{10}/n^{(11)}$	−3600	4320	15840	−86400	−2520	3240	−3960	−7200	8640	23760	−103680	20160	−25200
$2^21^8/n^{(10)}$	6300	−9900	−31680	194400	2520	−5040	8280	12600	−19800	−47520	233280	−25200	45360
$2^31^6/n^{(9)}$	−4500	10800	27720	−201600	−630	3150	−8190	−8820	21420	41580	−241920	10080	−35280
$2^41^4/n^{(8)}$	1170	−3670	−9900	94500	.	−630	3780	1890	−10710	−15120	113400	−630	10710
$2^51^2/n^{(7)}$	−90	1260	900	−16200	.	.	−630	.	1890	1890	−18900	.	−630
$2^6/n^{(6)}$	.	−90	.	900	.	.	.	.	.	.	.	.	.
$31^8/n^{(10)}$	1200	−1200	−5280	28800	840	−840	840	2400	−2400	−7920	34560	−6720	6720
$321^7/n^{(9)}$	−3600	4800	18480	−115200	−1260	2100	−2940	−7560	9960	27720	−138240	13440	−20160
$32^21^3/n^{(8)}$	2880	−6480	−19800	151200	210	−1470	3570	6720	−14280	−28980	181440	−5040	18480
$[32^21^3]/n^{(7)}$	−480	3360	6900	−72000	.	210	−1680	−1260	7980	8820	−88200	.	−5040
$32^41/n^{(6)}$	.	−480	−300	7200	.	.	210	.	−1260	−630	12600	.	.
$3^31^6/n^{(6)}$	540	−540	−2640	16800	140	−140	140	1120	−1120	−4200	20160	−1680	1680
$3^321^5/n^{(7)}$	−600	1140	4200	−36000	.	140	−280	−1680	2800	6720	−42000	560	−2240
$3^32^21^2/n^{(6)}$	30	−630	−1500	19800	.	.	140	210	−1890	−1260	21000	.	560
$3^32^4/n^{(5)}$	.	30	.	−600	.	.	.	.	210	.	−2100	.	.
$3^31^6/n^{(6)}$	40	−40	−200	2400	.	.	.	140	−140	−560	2800	.	.
$3^321/n^{(5)}$	.	40	100	−2400	.	.	.	.	140	.	−1400	.	.
$3^6/n^{(4)}$	.	.	.	100	.	.	.	.	.	.	.	.	.
$41^8/n^{(9)}$	−300	300	1320	−7200	−210	210	−210	−420	420	1980	−8640	1680	−1680
$[421^6]/n^{(6)}$	810	−1110	−3960	25200	210	−420	630	1050	−1470	−6300	30240	−2520	4200
$42^21^5/n^{(7)}$	−540	1350	3150	−27000	.	210	−630	−630	1680	5670	−31500	420	−2940
$42^31^3/n^{(6)}$	75	−615	−600	9900	.	.	210	.	−630	−1260	9450	.	420
$42^41/n^{(5)}$	.	75	.	−900	.	.	.	.	.	.	.	.	.
$431^7/n^{(7)}$	−240	240	1200	−7200	−35	35	−35	−280	280	1890	−9240	560	−560
$4321^5/n^{(6)}$	180	−420	−1200	10800	.	−35	70	315	−595	−2520	14700	.	560
$432^21/n^{(5)}$	.	180	150	−3600	.	.	−35	.	315	315	−4200	.	.
$43^31^3/n^{(5)}$	−10	10	50	−600	.	.	.	−35	35	280	−1400	.	.
$43^22/n^{(4)}$	.	−10	.	300	.	.	.	.	−35	.	350	.	.
$4^21^4/n^{(4)}$	30	−30	−150	900	.	.	.	.	.	−210	1050	−35	35
$[4^221^2]/n^{(5)}$	−15	45	75	−900	.	.	.	.	.	210	−1050	.	−35
$4^23^2/n^{(4)}$	.	−15	.	225	.	.	.	.	.	.	.	.	.
$4^231/n^{(4)}$	.	.	.	.	.	.	.	.	.	−35	175	.	.
$4^41/n^{(4)}$	.	.	.	.	.	.	.	.	.	.	.	.	.
$51^7/n^{(6)}$	36	−36	−264	1440	42	−42	42	84	−84	−252	1728	−336	336
$521^5/n^{(7)}$	−72	108	720	−4320	−21	63	−105	−168	252	630	−5544	336	−672
$52^21^3/n^{(6)}$	18	−90	−450	3240	.	−21	84	63	−231	−378	5040	.	336
$52^31/n^{(5)}$	.	18	30	−360	.	.	−21	.	63	63	−1260	.	.
$531^5/n^{(6)}$	24	−24	−240	1440	.	.	.	42	−42	−168	1680	−56	56
$5321^3/n^{(5)}$	.	24	180	−1440	.	.	.	−21	63	84	−2100	.	−56
$[532^21]/n^{(5)}$	.	.	.	.	.	.	.	.	−21	.	420	.	.
$53^21/n^{(4)}$	.	.	−10	120	.	.	.	.	.	.	140	.	.
$541^5/n^{(5)}$	−6	6	60	−360	.	.	.	.	.	42	−420	.	.
$5421/n^{(4)}$	.	−6	−15	180	.	.	.	.	.	−21	315	.	.
$543/n^{(3)}$	.	.	.	.	.	.	.	.	.	.	−35	.	.
$5^21^6/n^{(4)}$	.	.	−6	36	.	.	.	.	.	.	42	.	.
$5^22/n^{(3)}$	.	.	.	.	.	.	.	.	.	.	−21	.	.
$61^6/n^{(7)}$	−6	6	24	−240	−7	7	−7	−14	14	42	−168	56	−56
$621^4/n^{(6)}$	12	−18	−60	720	.	−7	14	21	−35	−84	420	−28	84
$62^21^2/n^{(5)}$	−3	15	30	−540	.	.	−7	.	21	21	−210	.	−28
$[62^2]/n^{(4)}$	.	−3	.	60	.	.	.	.	.	.	.	.	.
$631^3/n^{(5)}$	−4	4	20	−240	.	.	.	−7	7	28	−140	.	.
$6321/n^{(4)}$	.	−4	−10	240	.	.	.	.	−7	.	70	.	.
$63^2/n^{(3)}$	.	.	.	−20	.	.	.	.	.	.	.	.	.
$641^4/n^{(4)}$	*1*	−1	−5	60	.	.	.	.	.	−7	35	.	.
$642/n^{(3)}$	1	*1*	.	−30	.	.	.	.	.	.	.	.	.
$651/n^{(3)}$	5	.	*1*	−12	.	.	.	.	.	.	−7	.	.
$6^2/n^{(3)}$	30	30	12	*1*	.	.	.	.	.	.	.	.	.
$71^6/n^{(6)}$	.	.	.	.	*1*	−1	1	2	−2	−6	24	−8	8
$721^3/n^{(5)}$	.	.	.	.	1	*1*	−2	−3	5	12	−60	.	−8
$[72^21]/n^{(4)}$	.	.	.	.	1	2	*1*	.	−3	−3	30	.	.
$731^3/n^{(4)}$	.	.	.	.	1	3	.	*1*	−1	−4	20	.	.
$732/n^{(3)}$	.	.	.	.	1	4	3	1	*1*	.	−10	.	.
$741/n^{(3)}$	7	.	.	.	1	6	3	4	.	*1*	−5	.	.
$75/n^{(2)}$	35	.	7	.	1	10	15	10	10	5	*1*	.	.
$81^6/n^{(5)}$	.	.	.	.	8	.	.	.	.	.	.	*1*	−1
$821^3/n^{(4)}$	.	.	.	.	8	8	.	.	.	.	.	1	*1*
$82^2/n^{(3)}$	.	.	.	.	8	16	8	.	.	.	.	1	2
$831/n^{(3)}$	.	.	.	.	8	24	.	8	.	.	.	1	3
$84/n^{(2)}$	28	28	.	.	8	48	24	32	.	8	.	1	6
$[91^2]/n^{(4)}$	.	.	.	.	36	36	.	.	.	.	.	9	.
$921/n^{(3)}$	.	.	.	.	36	72	36	.	.	.	.	9	9
$93/n^{(2)}$	.	.	.	.	36	144	108	36	36	.	.	9	27
$10, 1^2/n^{(3)}$	210	.	.	.	120	360	.	120	.	.	.	45	45
$10, 2/n^{(2)}$	210	210	.	.	120	480	360	120	120	.	.	45	90
$11, 1/n^{(2)}$	2310	.	462	.	330	1980	990	1320	.	330	.	165	495
$12/n$	13860	13860	5544	462	792	7920	11880	7920	7920	3960	792	495	2970

Weight 12 (vi)	h_{61^2}	h_{521}	h_{44}	h_{51^2}	h_{431}	h_{422}	$h_{50,1^2}$	$h_{10,2}$	$h_{11,1}$	h_{12}
$[1^{12}]/n^{(12)}$	−5040	−10080	30240	40320	−40320	80640	−362880	362880	36288000	−39916800
$[21^{10}]/n^{(11)}$	30240	55440	−181440	−181440	221760	−483840	1814400	−2177280	−19958400	239500800
$[2^21^6]/n^{(10)}$	−70560	−110880	408240	272160	−453600	1088640	−3175200	4989600	39916800	−538876800
$[2^31^4]/n^{(9)}$	80640	95760	−423360	−151200	423360	−1118880	2268000	−5443200	−34927200	558835200
$[2^41^2]/n^{(8)}$	−45990	−31500	200340	22680	−173880	498960	−567000	2835000	12474000	−261954000
$[2^51^2]/n^{(7)}$	11340	1890	−37800	.	22680	−68040	22680	−589680	−1247400	44906400
$[2^6]/n^{(6)}$	−630	.	1890	.	.	.	.	22680	.	−1247400
$[31^9]/n^{(10)}$	−6720	−18480	60480	60480	−60480	161280	−604800	604800	6652800	−79833600
$[321^7]/n^{(9)}$	26880	67200	−241920	−151200	211680	−665280	1814400	−2419200	−23284800	319334400
$[32^21^3]/n^{(8)}$	−38640	−75600	312480	90720	−241920	907200	−1512000	3326400	24948000	−419126400
$[32^31]/n^{(7)}$	23520	25200	−141120	−7560	98280	−438480	302400	−1814400	−8316000	199584000
$32^41/n^{(6)}$	−5040	−630	17640	.	−7560	45360	.	302400	415800	−24948000
$3^21^6/n^{(6)}$	−1680	−10080	36960	20160	−20160	100800	−252000	252000	3326400	−46569600
$3^221^4/n^{(7)}$	3920	19600	−77280	−15120	35280	−241920	302400	−554400	−5544000	99792000
$3^22^21^2/n^{(6)}$	−2800	−6720	31920	.	−15120	136080	−37800	340200	1663200	−49896000
$3^22^3/n^{(5)}$	560	.	−1680	.	.	−7560	.	−37800	.	3326400
$3^31^3/n^{(6)}$	.	−1680	6720	560	−560	21280	−16800	16800	369600	−7392000
$3^321/n^{(5)}$	.	560	−2240	.	560	−16800	.	−16800	−92400	4435200
$3^3/n^{(4)}$	.	.	.	.	.	560	.	.	.	−92400
$41^8/n^{(8)}$	1680	3360	−15120	−15120	15120	−30240	151200	−151200	−1663200	19958400
$[421^6]/n^{(8)}$	−3880	−10080	55440	30240	−45360	105840	−378000	529200	4989600	−69854400
$42^21^4/n^{(7)}$	7140	8400	−63000	−11340	41580	−113400	226800	−604800	−4158000	74844000
$42^31^2/n^{(6)}$	−3360	−1260	22680	.	−11340	34020	−18900	245700	831600	−24948000
$42^4/n^{(5)}$	420	.	−1890	.	.	.	.	−18900	.	1247400
$431^5/n^{(7)}$	560	2800	−16800	−7560	7560	−30240	100800	−100800	−1386000	19958400
$4321^3/n^{(6)}$	−1120	−4200	30240	2520	−10080	57960	−75600	176400	1663200	−33264000
$432^21/n^{(5)}$	560	420	−8400	.	2520	−18900	.	−75600	−207900	9979200
$43^31^2/n^{(5)}$	.	560	−3920	.	.	−7560	4200	−4200	−138600	3326400
$43^22/n^{(4)}$	.	.	560	.	.	2520	.	4200	.	−831600
$4^31^4/n^{(6)}$	−35	−70	1890	630	−630	1260	−9450	9450	138600	−2079000
$[4^221^2]/n^{(6)}$	70	105	−2940	.	630	−1890	3150	−12600	−103950	2494800
$4^32^2/n^{(4)}$	−35	.	525	.	.	.	.	3150	.	−311850
$4^321/n^{(4)}$	.	−35	700	.	.	630	.	.	11550	−415800
$4^4/n^{(3)}$	.	.	−35	.	.	.	.	.	.	11550
$51^7/n^{(8)}$	−336	−672	2016	3024	−3024	6048	−30240	30240	332640	−3991680
$521^5/n^{(7)}$	1008	1680	−6048	−4536	7560	−18144	60480	−90720	−831600	11975040
$52^21^3/n^{(6)}$	−1008	−1008	5040	756	−5292	15120	−22680	83160	498960	−9979200
$52^31/n^{(5)}$	336	.	−1008	.	756	−2268	.	−22680	−41580	1995840
$531^4/n^{(6)}$	−56	−448	1680	1008	−1008	5040	−15120	15120	221760	−3326400
$5321^2/n^{(5)}$	112	504	−2016	.	1008	−7560	5040	−20160	−166320	3991680
$532^3/n^{(4)}$	−56	.	168	.	.	756	.	5040	.	−498960
$53^21/n^{(4)}$	.	−56	224	.	.	1008	.	.	9240	−332640
$541^3/n^{(5)}$	.	.	−336	−126	126	−252	2520	−2520	−41580	665280
$5421/n^{(4)}$	.	.	336	.	−126	378	.	2520	13860	−498960
$543/n^{(3)}$	.	.	−56	.	.	−126	.	.	.	55440
$5^21^4/n^{(4)}$	.	.	.	.	.	.	−126	126	2772	−40896
$5^22/n^{(3)}$	.	.	.	.	.	.	.	−126	.	16632
$61^6/n^{(7)}$	56	112	−336	−504	504	−1008	5040	−5040	−55440	665280
$621^4/n^{(6)}$	−140	−224	840	504	−1008	2520	−7560	12600	110880	−1663200
$62^21^2/n^{(5)}$	112	84	−504	.	504	−1512	1260	−8820	−41580	997920
$62^3/n^{(4)}$	−28	.	84	.	.	.	.	1260	.	−83160
$631^2/n^{(4)}$	.	56	−224	−84	84	−672	1680	−1680	−27720	443520
$6321/n^{(4)}$	.	−28	112	.	−84	756	.	1680	9240	−332640
$63^2/n^{(3)}$	.	.	.	.	.	−84	.	.	.	18480
$641^2/n^{(4)}$	.	.	56	.	.	.	−210	210	4620	−83160
$642/n^{(3)}$	.	.	−28	.	.	.	.	−210	.	27720
$651/n^{(3)}$	.	.	.	.	.	.	.	.	−462	11088
$6^2/n^{(3)}$	.	.	.	.	.	.	.	.	.	−462
$71^5/n^{(6)}$	−8	−16	48	72	−72	144	−720	720	7920	−95040
$721^3/n^{(6)}$	16	24	−96	−36	108	−288	720	−1440	−11880	190080
$[72^21]/n^{(4)}$	−8	.	24	.	−36	108	.	720	1080	−71280
$731^2/n^{(6)}$	.	−8	32	.	.	72	−120	120	2640	−47520
$732/n^{(3)}$	.	.	.	.	.	−36	.	−120	.	15840
$741/n^{(3)}$	.	.	−8	.	.	.	.	.	−330	7920
$75/n^{(3)}$	.	.	.	.	.	.	.	.	.	−792
$81^6/n^{(8)}$	1	2	−6	−9	9	−18	90	−90	−990	11880
$821^4/n^{(6)}$	−2	−3	12	.	−9	27	−45	135	990	−17820
$82^21^2/n^{(5)}$	1	.	−3	.	.	.	.	−45	.	2970
$831/n^{(3)}$	.	1	−4	.	.	−9	.	.	−165	3960
$[84]/n^{(3)}$	3	4	1	.	.	.	.	.	.	−495
$[91^7]/n^{(4)}$	.	.	.	1	−1	2	−10	10	110	−1320
$921/n^{(3)}$	.	.	.	1	1	−3	.	−10	−55	1320
$[93]/n^{(3)}$	.	9	.	1	3	1	.	.	.	−220
$10,1^2/n^{(3)}$	.	.	.	10	.	.	1	−1	−11	132
$10,2/n^{(3)}$	45	.	.	10	10	.	1	1	.	−66
$11,1/n^{(3)}$	.	165	.	55	55	.	11	.	1	−12
$[12]/n$	1485	1980	495	220	660	220	66	66	12	1

An angular transformation for the serial correlation coefficient

By G. M. JENKINS

University College, London

1. Introduction and summary

Quenouille (1948) has investigated the effect of applying Fisher's z-transformation to the distribution of the serial correlation coefficient of lag 1 in a circularly correlated universe. Apart from the fact that when ρ (the correlation in the parent population) is zero, the distributions of the serial correlation coefficient and the ordinary correlation coefficient have the same form, there does not seem to be any *a priori* reason for adopting this procedure. In fact, when $\rho \neq 0$ the distributions are quite distinct.

He deduces that 'compared with the transformation of the ordinary correlation coefficient, this transformation has the disadvantage that the variance of $z = \tanh^{-1} r$ is to the first order dependent on the value of ρ. Furthermore, for $|\rho|$ large, the mean value of z deviates widely from $\zeta = \tanh^{-1}\rho$.'

It is emphasized that the distribution is the appropriate one in the linear Markoff case only, viz. when

$$x_t = \rho x_{t-1} + \epsilon_t \quad \text{and} \quad \epsilon_t \text{ is a } N(0, \sigma) \text{ variate.} \tag{1}$$

Hence a transformation and a test of significance based thereon can only be applied once it is established that (1) is the correct mathematical model for the situation.

An inverse sine transformation is proposed* which is designed to stabilize the variance, and it is found that the limiting distribution of the transformed variate is approximately normal. The mean value is still sensitive to fluctuations in ρ as the latter tends to unity, but this is not so great as in the $\tanh^{-1} r$ transformation. This raises no serious practical problems, and it is shown that a simple test may be applied for values of ρ as large as 0·9.

Approximate confidence intervals are suggested which are in close agreement with those obtained by Quenouille (1949) using a different argument.

2. Derivation of the transformation

Leipnik (1947) derives the following distribution for the serial correlation coefficient of lag 1 calculated from a sample of n:

$$h(r) = \frac{\Gamma(\frac{1}{2}n+1)}{\Gamma(\frac{1}{2}n+\frac{1}{2})\,\Gamma(\frac{1}{2})}(1-r^2)^{\frac{1}{2}(n-1)}(1+\rho^2-2\rho r)^{-\frac{1}{2}n}, \tag{2}$$

where

$$r = \frac{x_1x_2+x_2x_3+\ldots+x_nx_1}{x_1^2+x_2^2+\ldots+x_n^2},$$

and the x's are independently and normally distributed about zero.

The mean and variance of r are given by

$$\mathscr{E}(r) = \alpha = \frac{\rho n}{n+2}, \tag{3}$$

and

$$\sigma_r^2 = \frac{1}{n+2}\left[1 - \frac{\rho^2 n(n+1)}{(n+2)(n+4)}\right] \tag{4}$$

$$= \frac{1}{n+2}[1-\lambda\alpha^2],$$

where

$$\lambda = \frac{(n+2)(n+1)}{n(n+4)} \doteqdot 1.$$

Hence

$$\sigma_r^2 \doteqdot \frac{1}{n+2}(1-\alpha^2). \tag{5}$$

---

* Since writing this paper, the author has learned that this transformation was suggested by F. Chartier (*Publications de L'Institut de Statistique de L'Université de Paris*, Vol. 1—Fascicule IV—1952) and also by Prof. M. G. Kendall in the course of a discussion on Prof. Hotelling's paper read to the Research Section of the Royal Statistical Society in May 1953.

We seek a transformation $y = f(r)$ such that the variance of y is approximately independent of ρ:

$$\operatorname{var} y \doteqdot \operatorname{var} r \left(\frac{dy}{dr}\right)^2_{r=\alpha},$$

and if the variance of y is to be a function of n and not of ρ we have

$$f(\alpha) = k \int \frac{d\alpha}{(1-\alpha^2)^{\frac{1}{2}}} = k \sin^{-1}\alpha,$$

whence $$r = \sin y/k.$$

Since the constant has no effect on the form of distribution of y, we may take

$$r = \sin y. \tag{6}$$

3. The limiting distribution of the transformed variate

Let $$r = \sin y, \quad \rho = \sin\lambda, \quad y - \lambda = x. \tag{7}$$

For large n it may be shown that $r - \rho = O(n^{-\frac{1}{2}})$. Therefore, since $\sin\theta \doteqdot \theta$ when θ is small,

$$y - \lambda = x = O(n^{-\frac{1}{2}}). \tag{8}$$

We apply the transformation (7) to $h(r)$ as given in (2) and obtain

$$h(x) = \frac{\Gamma(\frac{1}{2}n+1)}{\Gamma(\frac{1}{2}n+\frac{1}{2})\,\Gamma(\frac{1}{2})} \frac{\cos^n x \left[1 - \dfrac{\rho}{(1-\rho^2)^{\frac{1}{2}}}\tan x\right]^n}{\left\{1 + \dfrac{2\rho^2 - 2\rho^2\cos x - 2\rho(1-\rho^2)^{\frac{1}{2}}\sin x}{1-\rho^2}\right\}^{\frac{1}{2}n}}. \tag{9}$$

We need the following expansions:

$$\log_e\left\{\frac{\Gamma(\frac{1}{2}n+1)}{\Gamma(\frac{1}{2})\,\Gamma(\frac{1}{2}n+\frac{1}{2})}\right\} = \tfrac{1}{2}\log_e\frac{n}{2\pi} + \frac{1}{4n} + O\left(\frac{1}{n^2}\right), \tag{10a}$$

$$\log_e\{\cos x\} = -\tfrac{1}{2}x^2 - \tfrac{1}{12}x^4 + O\left(\frac{1}{n^3}\right), \tag{10b}$$

$$\log_e\{1 - \theta\tan x\} = -\theta(x + \tfrac{1}{2}x^3 + \tfrac{2}{15}x^5) - \tfrac{1}{2}\theta^2(x^2 + \tfrac{2}{3}x^4) - \tfrac{1}{3}\theta^3(x^3 + x^5) - \tfrac{1}{4}\theta^4 x^4 - \tfrac{1}{5}\theta^5 x^5 + O\left(\frac{1}{n^3}\right). \tag{10c}$$

In a similar manner we may expand

$$\log_e\left\{1 + \frac{2\rho^2 - 2\rho^2\cos x - 2\rho(1-\rho^2)^{\frac{1}{2}}\sin x}{1-\rho^2}\right\},$$

neglecting terms of higher order than $n^{-\frac{1}{2}}$.

Taking logarithms of both sides of (9) we obtain, after some manipulation,

$$\log_e\{h(x)\} = \tfrac{1}{2}\log_e\frac{n}{2\pi} - \tfrac{1}{2}nx^2 - \frac{1}{2}\frac{\rho}{(1-\rho^2)^{\frac{1}{2}}}nx^3 + \frac{1}{4n} - \frac{1}{24}\frac{(2+13\rho^2)}{(1-\rho^2)}nx^4 - \frac{1}{8}\frac{\rho}{(1-\rho^2)^{\frac{1}{2}}}(1+5\rho^2)nx^5 + O\left(\frac{1}{n^2}\right),$$

from which it follows that

$$h(x) = \sqrt{\frac{n}{2\pi}}\, e^{-\frac{1}{2}nx^2}\left\{1 - \frac{n\rho}{2(1-\rho^2)^{\frac{1}{2}}}x^3 + \frac{1}{4n} - \frac{1}{24}\frac{(2+13\rho^2)}{1-\rho^2}nx^4 + \frac{1}{8}\frac{\rho^2}{1-\rho^2}n^2x^6 - \frac{1}{8}\frac{\rho(1+5\rho^2)}{(1-\rho^2)^{\frac{1}{2}}}nx^5 - \frac{1}{8}\frac{\rho}{(1-\rho^2)^{\frac{1}{2}}}x^3 + \frac{1}{48}\frac{\rho(2+13\rho^2)}{(1-\rho^2)^{\frac{1}{2}}}n^2x^7 - \frac{1}{48}\frac{\rho^2}{(1-\rho^2)^{\frac{1}{2}}}n^2x^9 + O\left(\frac{1}{n^2}\right)\right\}. \tag{11}$$

4. Moments and coefficients of skewness and kurtosis of $h(x)$

From equation (11) we obtain the moments, as follows:

$$\mu_1' = -\frac{3}{2}\frac{\rho}{(1-\rho^2)^{\frac{1}{2}}}\frac{1}{n}+\frac{1}{8}\frac{\rho}{(1-\rho^2)^{\frac{1}{2}}}(17-2\rho^2)\frac{1}{n^2}+O\left(\frac{1}{n^3}\right), \tag{12}$$

$$\mu_2 = \frac{1}{n}-\frac{1}{2n^2}\frac{(2-5\rho^2)}{(1-\rho^2)}+O\left(\frac{1}{n^3}\right), \tag{13}$$

$$\mu_3 = -\frac{3\rho}{(1-\rho^2)^{\frac{1}{2}}}\frac{1}{n^2}+O\left(\frac{1}{n^3}\right), \tag{14}$$

$$\mu_4 = \frac{3}{n^2}-\frac{1}{n^3}\frac{(8-29\rho^2)}{(1-\rho^2)}+O\left(\frac{1}{n^4}\right), \tag{15}$$

$$\gamma_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = -\frac{3\rho}{n^{\frac{1}{2}}(1-\rho^2)^{\frac{1}{2}}}\left\{1+\frac{3}{4n}\frac{(2-5\rho^2)}{(1-\rho^2)}\right\}+O\left(\frac{1}{n^{\frac{5}{2}}}\right), \tag{16}$$

$$\gamma_2 = \frac{\mu_4}{\mu_2^2}-3 = \frac{2}{n}\frac{(7\rho^2-1)}{1-\rho^2}+O\left(\frac{1}{n^2}\right). \tag{17}$$

These results may be compared with those of Quenouille (1948, p. 262) for the $\tanh^{-1}$ transformation. The most noticeable difference is that in his case the second moment is not independent of ρ to the first order. Further, his series for the mean commences with a term containing the factor $(1-\rho^2)^{-1}$ instead of $(1-\rho^2)^{-\frac{1}{2}}$ as in equation (12), so that the mean value of $z = \tanh^{-1} r$ will diverge more from $\zeta = \tanh^{-1}\rho$ when $|\rho|$ is large than will y from λ.

5. Tests of significance

We assume that the theoretical model for the situation may be represented by the linear Markoff process $x_t = \rho x_{t-1}+\epsilon_t$, where ϵ_t is a $N(0,\sigma)$ variable. It is desired to test the hypothesis that the population correlation equals some value ρ.

The transformation $r = \sin y, \rho = \sin\lambda$ is such that we may take $y-\lambda$ to be approximately normally distributed with mean and variance given by (12) and (13). The variance of y is independent of ρ to the first order.

If we let

$$\mu_1' = \frac{\beta(\rho)}{n}+\frac{\gamma(\rho)}{n^2}+O\left(\frac{1}{n^3}\right), \tag{18}$$

$$\mu_2 = \frac{1}{n}+\frac{\delta(\rho)}{n^2}+O\left(\frac{1}{n^3}\right), \tag{19}$$

the stability of the mean and variance of y is illustrated by the following table:

ρ	$\beta(\rho)$	$\gamma(\rho)$	$\delta(\rho)$
0·0	0·00	0·00	−1·00
0·2	−0·31	0·43	−0·94
0·5	−0·87	1·59	−0·50
0·8	−2·00	7·27	+1·66
0·9	−3·10	20·87	+5·39

Suppose that we take $n = 35$. It may be seen that even for $\rho = 0{\cdot}5$ the disturbances are negligible and tend to annul one another since they are opposite in sign.

For values of ρ up to 0·9 the corrections are still small (being majorized by the first-order term), but they would have to be allowed for in a test of significance. There is, however, a large increase in the value of $\gamma(\rho)$ for values of ρ between 0·9 and 1·0, so that it is highly unlikely that the transformation can be used at all in this range.

This raises no serious practical problems, since in the case of a single series it will not often be required to test whether the population value of ρ is equal to some number exceeding 0·9. It is, however, possible that the equality of values of ρ in two series, each with high serial correlation, may be the subject of a test. It appears that the transformation would not be suitable in such a case.

The variance as given by (13) is very stable, e.g. it alters from $1/n - 1/n^2$ to $1/n - 1/(2n^2)$ in the range 0·0–0·5. These are approximately given by $1/(n+1)$ and $1/(n+2)$ respectively. We may thus take $y - \lambda$ to be approximately normally distributed with zero mean and variance $1/(n+1)$ (provided that ρ is small and n is reasonably large).

If it is known that ρ lies in the range 0·0 to 0·5, then an average correction may be applied to the variance in the form $1/(n+1·5)$.

6. Approximate confidence intervals

Quenouille (1949) has obtained confidence intervals for ρ by considering a transformation of the distribution function of r into a form which can be expanded in terms of Student's 't'. This procedure was necessary since it would be difficult to construct confidence intervals for ρ using the $r = \tanh z$ transformation since the variance of z is dependent on ρ. These intervals are exact in the sense that they are derived without approximation from (2), but approximate in that (2) is only a smoothed form of the exact distribution of r. They are, however, asymmetrical with respect to probabilities, the degree of asymmetry being a function of ρ.

The inverse sine transformation results in confidence intervals which, although they are not exact since y is only approximately normally distributed, are symmetrical and are smaller on the average. When n is reasonably large, they are in close agreement with those obtained by Quenouille, viz.

$$\rho = r \pm t_{n+1,\alpha} \frac{(1-r^2)^{\frac{1}{2}}}{\sqrt{(n+1)}}. \tag{20}$$

Using the fact that $y - \lambda$ is approximately a $N(0, (n+1)^{-1})$ variate, we have that

$$P\{-Z_\alpha \leqslant \sqrt{(n+1)}\,(\sin^{-1} r - \sin^{-1}\rho) \leqslant Z_\alpha\} \doteqdot 1 - 2\alpha,$$

which results in approximate confidence intervals for ρ in the form

$$\rho = r\cos\frac{Z_\alpha}{\sqrt{(n+1)}} \pm (1-r^2)^{\frac{1}{2}} \sin\frac{Z_\alpha}{\sqrt{(n+1)}}. \tag{21}$$

Using a result given by Peiser (1945), viz. that

$$t_{n+1,\alpha} = Z_\alpha + \frac{(Z_\alpha^3 + Z_\alpha)}{4(n+1)} + O\left(\frac{1}{n^2}\right), \tag{22}$$

and assuming that n is large enough to take

$$\sin\frac{Z_\alpha}{\sqrt{(n+1)}} \doteqdot \frac{Z_\alpha}{\sqrt{(n+1)}}, \quad \cos\frac{Z_\alpha}{\sqrt{(n+1)}} \doteqdot 1,$$

(20) and (21) reduce to

$$\rho = r \pm \frac{Z_\alpha (1-r^2)^{\frac{1}{2}}}{\sqrt{(n+1)}}.$$

This result lends support to the assumption of normality of y.

If we define L_1 and L_2 to be interval lengths in (20) and (21), we obtain

$$L_1 = 2t_{n+1,\alpha}\frac{(1-r^2)^{\frac{1}{2}}}{\sqrt{(n+1)}}, \quad L_2 = 2(1-r^2)^{\frac{1}{2}}\frac{\sin Z_\alpha}{\sqrt{(n+1)}}.$$

Let

$$\mathscr{E}\{(1-r^2)^{\frac{1}{2}}\} = \phi(\rho, n),$$

Then

$$\mathscr{E}(L_2) = \phi(\rho,n)\frac{\sin Z_\alpha}{\sqrt{(n+1)}} \leqslant \phi(\rho,n)\frac{Z_\alpha}{\sqrt{(n+1)}} < \phi(\rho,n)\frac{t_\alpha}{\sqrt{(n+1)}} = \mathscr{E}(L_1).$$

7. Experimental results

Kendall (1949) has constructed artificial linear autoregressive series of the form $x_t = \rho x_{t-1} + \epsilon_t$, where ϵ_t is a $N(0, 1)$ variate. His Series 9 (for which $\rho = 0·7$) was divided into ten subseries, each containing fifty members. Circular serial correlation coefficients of lag 1 were computed for the whole series and for each of the ten subseries, the values of x being taken about the theoretical mean zero.

(*a*) 95 % confidence intervals for ρ were obtained from the eleven values of r, using Quenouille's method (I) and that based on the $\sin^{-1}$ transformation (II).

(*b*) Estimates of the mean value and variance of $y = \sin^{-1} r$ were obtained from the ten values of y from the subseries and compared with the theoretical values calculated from (12) and (13).

The results derived for (*a*) were as follows:

Sub-series	r	y	95 % confidence limits	
			Method I	Method II
1	0·544	0·5752	0·308–0·780	0·296–0·751
2	0·644	0·6997	0·430–0·858	0·412–0·827
3	0·790	0·9110	0·618–0·962	0·595–0·927
4	0·651	0·7089	0·438–0·864	0·421–0·832
5	0·683	0·7519	0·481–0·885	0·459–0·855
6	0·618	0·6662	0·397–0 839	0·382–0·808
7	0·769	0·8773	0·589–0·949	0·567–0·914
8	0·618	0·6662	0·397–0·839	0·382–0·808
9	0·578	0·6163	0·349–0·807	0·335–0·778
10	0·784	0·9011	0·610–0·958	0·587–0·923
Whole series	0·697	0·7712	0·634–0·760	0·632–0·757

It will be seen that for the whole series, Quenouille's method resulted in the confidence limits (0·634, 0·760) for ρ, whilst the $\sin^{-1}$ transformation yielded limits (0·632, 0·757). The agreement is close, as might be expected for a series containing 500 members.

In the case of those based on the subseries there is a distinct displacement of the confidence interval as a result of using the $\sin^{-1}$ transformation. This shift is very nearly the same for all ten intervals, being on the average, 0·032 for the upper limit and 0·018 for the lower. This was to be expected since the intervals obtained by method I were asymmetrical with respect to probabilities—the translation results in biased intervals which are symmetrical. It is also noted that each interval length in method II is less than the corresponding interval obtained by method I, the average interval lengths being 0·400 and 0·412 respectively.

The results for comparison (*b*) were as follows:

	Expected value	Calculated value
Mean	0·743	0·737
Variance	0·0203	0·0144

The expected and calculated values of the mean agree very closely, whilst a χ^2-test of significance showed that the calculated value of the variance was in reasonable accord with the expected value as obtained from equation (13).

In conclusion, I wish to express my warm thanks to Dr N. L. Johnson for his aid during the preparation of this work.

REFERENCES

KENDALL, M. G. (1949). *Biometrika*, **36**, 267.
LEIPNIK, R. B. (1947). *Ann. Math. Statist.* **18**, 80.
PEISER, A. M. (1945). *Ann. Math. Statist.* **14**, 56.
QUENOUILLE, M. H. (1948). *Biometrika*, **35**, 261.
QUENOUILLE, M. H. (1949). *Proc. Camb. Phil. Soc.* **45**, 483.

A distribution arising in the study of infectious diseases

By J. O. IRWIN

Statistical Research Unit of the Medical Research Council, London School of Hygiene

The late Prof. Greenwood was for many years interested in the infectiousness of measles. During his study of the distribution of numbers of cases of measles in households he developed the chain-binomial distribution (Greenwood, 1931). This appeared to give a good fit to the distribution of the number of secondary cases. However, further analysis by E. B. Wilson and his colleagues (1938, 1947) and by Greenwood himself (1949) showed that chain-binomials were not satisfactory. Greenwood then considered another model.

Suppose that n children are exposed to risk in a household and each child has a probability p of being attacked, but that these probabilities are not independent but correlated. The simplest assumption is that all these correlations are equal, say r. Greenwood saw that the variance of the number attacked would be $npq\{1+(n-1)r\}$. This result is easily obtained, for if

$$\begin{aligned} x_u &= 1 \text{ if } u\text{th child is attacked,} \\ &= 0 \text{ if } u\text{th child is not attacked,} \\ \sigma_u^2 &= pq \text{ for all } u, \end{aligned}$$

we require the variance of $X = \sum_{u=1}^{n} x_u$,

and this is
$$\sum_{u=1}^{n} \sigma_u^2 + \sum_{u \neq v} r\sigma_u\sigma_v = npq + n(n-1)rpq$$
$$= npq(1+\overline{n-1}r).$$

He asked me whether one could obtain the frequency distribution of the number attacked ? The following solution of the problem is simpler than the one I originally obtained.

Let $P_1(k \mid r)$ be the probability that when $x_u = 1$, out of r other children k will be attacked. Let $P_0(k \mid r)$ be the probability of the same occurrence when $x_u = 0$. Let $P(k, n)$ be the probability that k out of n are attacked. Then
$$P(k,n) = pP_1(k-1 \mid n-1) + qP_0(k \mid n-1). \tag{1}$$

Thus if we can obtain $P_1(0 \mid 1)$, $P_0(0 \mid 1)$, $P_1(1 \mid 1)$, $P_0(1 \mid 1)$ the distribution can be built up.

Consider x_1 and x_2, and suppose that their joint frequency distribution is

		x_2 = 0	1	
x_1	0	p_{00}	p_{01}	q
	1	p_{10}	p_{11}	p
		q	p	1

The regression of x_2 on x_1 is r, for the two standard deviations are equal. When $x_1 = 0$ the mean value of x_2 is p_{01}/q, when $x_1 = 1$ it is p_{11}/p. Thus

$$(p_{11}/p) - (p_{01}/q) = r$$

and
$$p_{11} + p_{01} = p.$$

Thus
$$p_{01} = p_{10} = p(1-r), \quad p_{00} = q^2 + pqr, \quad p_{11} = p^2 + pqr,$$

and the frequency distribution is

		x_1 = 0	1	
x_2	0	q^2+pqr	$pq(1-r)$	q
	1	$pq(1-r)$	p^2+pqr	p
		q	p	1

(2)

Thus
$$\left.\begin{aligned} P_1(1 \mid 1) &= p + qr, \\ P_0(1 \mid 1) &= p(1-r), \\ P_1(0 \mid 1) &= q(1-r), \\ P_0(0 \mid 1) &= q + pr. \end{aligned}\right\} \tag{3}$$

Now consider the ordinary hypergeometric distribution, whose generating function is

$$\frac{(Nq)_{(n)}}{N_{(n)}} F(-n, -Np, Nq-n+1, A).$$

It satisfies the recurrence formula (1) and

$$\left.\begin{aligned} P_1(1 \mid 1) &= \frac{Np-1}{N-1}, \\ P_0(1 \mid 1) &= \frac{Np}{N-1}, \\ P_1(0 \mid 1) &= \frac{Nq}{N-1}, \\ P_0(0 \mid 1) &= \frac{Nq-1}{N-1}. \end{aligned}\right\} \tag{4}$$

(4) and (3) may be identified by taking $r = -\frac{1}{(N-1)}$, but r is positive, hence $N = -\left(\frac{1-r}{r}\right)$, which is *negative*. In the usual case of the hypergeometric distribution, N is a positive integer; here it is negative and not necessarily an integer. Thus the distribution is

$$\frac{(Nq)_{(n)}}{N_{(n)}} F(-n, -Np, Nq-n+1, A) \tag{5}$$

with $$N = -\left(\frac{1-r}{r}\right).$$

For example, for $n = 2$ we find

$$\frac{Nq(Nq-1)}{N(N-1)}\left\{1 + \frac{2NpA}{1!\,(Nq-1)} + \frac{Np(Np-1)A^2}{Nq(Nq-1)}\right\},$$

and this easily reduces to

$$q(q+rp)\left\{1 + \frac{2p(1-r)A}{(q+rp)} + \frac{p(1-r)(p+qr)A^2}{q(1-r)(q+rp)}\right\}, \tag{6}$$

and the three frequencies are

$$\begin{aligned} 0 \quad & q^2+pqr, \\ 1 \quad & 2pq(1-r), \\ 2 \quad & p^2+pqr, \end{aligned}$$

as is obvious from (2).

For $n = 3$ we find

$$\left.\begin{aligned} 0 \quad & q(q+rp)(q+rp+r)/(1+r), \\ 1 \quad & 3pq(q+rp)(1-r)/(1+r), \\ 2 \quad & 3pq(p+rq)(1-r)/(1+r), \\ 3 \quad & p(p+qr)(p+qr+r)/(1+r). \end{aligned}\right\} \tag{7}$$

The case when N is negative but an integer arises in the ordinary model of drawing balls from a bag, when after each ball is drawn it is replaced and an additional ball of the same colour inserted. However, the present case is much more general than that.

We might change our model somewhat taking a definite period T of exposure to risk and dividing it into n equal intervals (n large). We can suppose that a single case may occur in any interval with probability p, but that these probabilities are correlated with correlation r as before. Then the distribution (5) will hold. Now put $np = m$, $nr = \alpha$, and let $n \to \infty$, $p \to 0$, $r \to 0$, in such a way that m and α remain finite. In this case the generating function of the distribution becomes

$$(1+\alpha)^{-m/\alpha}\left(1 - \frac{\alpha A}{1+\alpha}\right)^{-m/\alpha},$$

or the negative binomial distribution. The simplest way to show this is to consider the factorial moments.

Greenwood (1949) fitted the distribution (7) to 100 sets of three cases (Wilson and Worcester's data). p was estimated by equating $3p$ to the observed mean and $npq(1+\overline{n-1}r)$ to the variance. The result is:

Obs.	Exp.	
4	4·44	$\left.\begin{matrix}\chi^2 = 0{\cdot}331\\ \nu = 1\\ P = 0{\cdot}57\end{matrix}\right\}$.
11	9·68	
18	19·32	
67	66·56	
100	100	

This is a good fit, but is the only example available so far of fitting actual data with this distribution. Naturally much more experience is needed before we can conclude that it is really satisfactory.

REFERENCES

GREENWOOD, M. (1931). On the statistical measure of infectiousness. *J. Hyg., Camb.*, **31**, 336–51.
GREENWOOD, M. (1949). The infectiousness of measles. *Biometrika*, **36**, 1–8.
WILSON, E. B. (1947). *Proc. Nat. Acad. Sci., Wash.*, **33**, 68–72.
WILSON, E. B., BENNETT, C., ALLEN, A. & WORCESTER, J. (1938). *Proc. Amer. Phil. Soc.* **80**, 359–76.

A note on contagious distributions

BY H. R. THOMPSON

Applied Mathematics Laboratory, D.S.I.R., Wellington, New Zealand

1. Contagious distributions have been applied to ecological problems connected with the spread and distribution of plants over an area, where we have initial or 'parent' plants, and others ('offspring') associated with them in such a way as to render the use of the simple Poisson distribution as a description of the spreading process invalid. Earlier work in ecology was based on the assumption of a random distribution of plants, but this has generally been found to be an erroneous assumption, due often to the associative properties of neighbouring plants. Random samples of quadrats showed that in comparison with Poisson expectation there was an excess of plots with no plants and with a large number, a property described as 'contagion'. Neyman (1939) developed distributions with this property in the fields of bacteriology and entomology, but they are equally applicable to ecology. Thomas (1949) invented the double Poisson distribution to describe non-random distributions of plants. This note discusses another method of obtaining contagious distributions, due to Darwin (1951), of wider applicability, and draws attention to some shortcomings of the two earlier methods. Since this was written, a paper by Skellam (1952) has appeared, in which similar aspects of the problem have been discussed, but as the methods used here are different it was considered worth-while to present them without modification.

2. The derivation of Darwin's model is more formal than that of Neyman, and as far as possible a notation consistent with his is employed. Assume a random distribution of parent plants or centres with mean m per unit area, an assumption consistent on the whole with practical experience when an area is first invaded by a species. Taking an arbitrary origin, and referring co-ordinates of plants, which we assume to be dimensionless, to this origin, then for a parent at (ξ, η) the co-ordinate distances $x-\xi, y-\eta$ of an offspring at (x, y) obey a probability distribution $f(x-\xi, y-\eta)\,dx\,dy$ of being in the infinitesimal rectangle, area $dx\,dy$, left-hand lower corner at (x, y). Then for the test region R (which will usually be a quadrat, of side h, say)

$$\Pr\{1 \text{ offspring in } R \mid \text{centre at } (\xi, \eta)\} = \int_R f(x-\xi, y-\eta)\,dx\,dy$$
$$= P(\xi, \eta) \text{ say.}$$

For a group, centre (ξ, η), with n offspring, the probability generating function (p.g.f.) of the number in R (assuming individuals of a group to act independently) is

$$(P(\xi, \eta)\,z + 1 - P(\xi, \eta))^n,$$

the average value of which, given the probability distribution $p(n)$ of the number n of offspring per group, is

$$\sum_n p(n)(Pz+1-P)^n = f(z) \text{ say.}$$

Multiplying by the probability law of ξ and η and integrating over the whole area with respect to ξ and η, Darwin obtains the p.g.f. for the number of offspring in R due to all centres, assuming all areas to contribute independently to the number in R.

Now let

$$\begin{aligned}\Pr(1 \text{ centre in } d\xi d\eta) &= m d\xi d\eta,\\ \Pr(\text{no centre in } d\xi d\eta) &= 1-m d\xi d\eta,\\ \Pr(>1 \text{ centre in } d\xi d\eta) &= o(d\xi d\eta).\end{aligned}$$

Therefore the contribution to the p.g.f. from the small area $d\xi d\eta$ at (ξ, η) is, by the ordinary multiplication law of probabilities,

$$1+m(f(z)-1)\,d\xi d\eta$$

(since the p.g.f. of the number in R when there is no centre at (ξ, η) is 1).

Taking logarithms to the first order in $d\xi d\eta$ log p.g.f. due to the small area $d\xi d\eta$ is

$$m(f(z)-1)\,d\xi d\eta,$$

and hence log p.g.f. for the whole area F is, since centres are assumed to act independently,

$$\int_F m(f(z)-1)\,d\xi d\eta.$$

In this, $p(n)$ and $P(\xi, \eta)$ are arbitrary. Actually, $f(x-\xi, y-\eta)$ should be specified and $P(\xi, \eta)$ calculated, but since $f(x-\xi, y-\eta)$ is also arbitrary, it is formally only necessary to specify $P(\xi, \eta)$. A valid assumption is that it is a circularly symmetric function $P(r)$, so that transforming to polar co-ordinates and integrating over θ,

$$\log \text{p.g.f.} = \int_0^\infty 2\pi m(f(z)-1)\,r\,dr, \tag{1}$$

in which $\int_0^\infty 2\pi P(r)\,r\,dr = 1$. (Note that this gives the distribution of the offspring alone. The p.g.f. for centres and offspring combined would be obtained by multiplying the above p.g.f., $\pi(z)$ say, by the Poisson p.g.f. $\exp\{m(z-1)\}$. Moments are most easily obtained from the component cumulant generating functions, which are additive.)

That the above derivation is entirely equivalent to Neyman's model is seen by putting $p(n) = e^{-\lambda}\lambda^n/n!$, $P(\xi, \eta) = A^{-1}$ inside a given finite area A and 0 outside, the values used by Neyman to obtain his Type A contagious distribution. Then

$$f(z)-1 = e^{\lambda(z-1)A}-1, \quad \log \pi(z) = mA[e^{\lambda(z-1)A}-1].$$

Darwin's model gives a much wider range of distributions, however, because functions $P(\xi, \eta)$ can be used in which dispersal is theoretically possible over the whole field. Such a function is the two-dimensional Gaussian $e^{-\frac{1}{2}r^2/\sigma^2}/(2\pi\sigma^2)$, which might be considered an adequate approximation for describing some dispersal mechanisms. The form of Neyman's derivation limits the possibilities for $P(\xi, \eta)$ to dispersal inside a finite and fairly small area, with none outside. For, putting A, the area of the part of the field in which $P(\xi, \eta) > 0$, equal to M, the total area of the field, in Neyman's equations, we are led to a limit of the form $\lim_{N\to\infty}\left(\frac{X}{N}\right)^N$ for the p.g.f., which equals zero.

3. Darwin used his model to obtain a new method of generating the negative binomial distribution. Take

$$P(r) = \frac{1}{2\pi\sigma^2}e^{-\frac{1}{2}r^2/\sigma^2}, \quad p(n) = (1-\mu)\,\mu^n \quad (n = 0, 1, \ldots).$$

Then integration of (1) gives

$$\pi(z) = (1-\beta(z-1))^{-\alpha},$$

where

$$\alpha = 2\pi\sigma^2 m, \quad \beta = \mu/2\pi\sigma^2(1-\mu).$$

In practice we would wish to use $e^{-\frac{1}{2}r^2/\sigma^2}/(2\pi\sigma^2)$ to represent the distribution of distances between the plants themselves, i.e. have

$$f(x-\xi, y-\eta) = e^{-\frac{1}{2}r^2/\sigma^2}/(2\pi\sigma^2) \quad (r^2 = (x-\xi)^2+(y-\eta)^2).$$

The negative binomial is therefore an approximation, as it is necessary for $P(\xi, \eta)$ to equal

$$e^{-\frac{1}{2}r^2/\sigma^2}/(2\pi\sigma^2)\,d\xi\,d\eta$$

($r^2 = \xi^2 + \eta^2$, assuming the origin to be at the centre of R) for all points in R. It would appear therefore that R must be fairly small in relation to σ^2 to give a reasonable approximation. In fact, the approximation is fairly good so long as h/σ is not greater than about $\frac{1}{2}$.

4. We proceed to obtain from Darwin's model some examples of contagious distributions by substituting arbitrary functions for $p(n)$ and $P(r)$ in (1). The differing degrees of contagion can be compared by means of the index of dispersion μ_2/μ_1', or in other words by comparing second moments when the means are adjusted to be equal. This method is used in preference to a consideration of individual probabilities of occurrence $(\pi_0, \pi_1, \pi_2, \ldots)$ as giving a better overall idea of the situation. The following four common distributions for $p(n)$ are used: Poisson $[e^{-\lambda}\lambda^n/n!]$, binomial $\left[\binom{N}{n} p^n q^{N-n}, n = 0, 1, \ldots, N\right]$, geometric $[(1-\mu)\mu^n]$, and negative binomial $[(1-\alpha)^2(n+1)\alpha^n]$. Further, for $P(r)$ we take the forms:

$$\frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}r^2/\sigma^2}, \quad \frac{1}{2\pi\sigma^2} e^{-r/\sigma}, \quad \frac{3}{\pi R^3}(R-r), \quad \frac{6}{\pi R^4} r(R-r), \quad \frac{1}{\pi R^2},$$

(the last three inside a circle of radius R, and zero outside). We then obtain by integration exact forms for the p.g.f.'s, but except for the example given above, none of the others gives rise to common distributions. They are mostly rather complicated.

Table 1. *Values of the index of dispersion* μ_2/μ_1' $(\mu_1' = m\lambda)$

$p(n)$	$E(n)=\lambda$	$P(r)$				
		(1) $\frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}r^2/\sigma^2}$	(2) $\frac{1}{2\pi\sigma^2} e^{-r/\sigma}$	(3) $\frac{3}{\pi R^3}(R-r)$	(4) $\frac{6}{\pi R^4} r(R-r)$	(5) $\frac{1}{\pi R^2}$
(1) $(1-\mu)\mu^n$	$\frac{\mu}{1-\mu}$	$1+\frac{\lambda}{2\pi\sigma^2}$	$1+\frac{\lambda}{4\pi\sigma^2}$	$1+\frac{3\lambda}{\pi R^2}$	$1+\frac{12\lambda}{5\pi R^2}$	$1+\frac{2\lambda}{\pi R^2}$
(2) $(1-\alpha)^2(n+1)\alpha^n$	$\frac{2\alpha}{1-\alpha}$	$1+\frac{3\lambda}{8\pi\sigma^2}$	$1+\frac{3\lambda}{16\pi\sigma^2}$	$1+\frac{9\lambda}{4\pi R^2}$	$1+\frac{9\lambda}{5\pi R^2}$	$1+\frac{3\lambda}{2\pi R^2}$
(3) $\frac{e^{-\lambda}\lambda^n}{n!}$	λ	$1+\frac{\lambda}{4\pi\sigma^2}$	$1+\frac{\lambda}{8\pi\sigma^2}$	$1+\frac{3\lambda}{2\pi R^2}$ †	$1+\frac{6\lambda}{5\pi R^2}$	$1+\frac{\lambda}{\pi R^2}$ *
(4) $\binom{N}{n} p^n q^{N-n}$	Np	$1+\frac{\lambda-p}{4\pi\sigma^2}$	$1+\frac{\lambda-p}{8\pi\sigma^2}$	$1+\frac{3(\lambda-p)}{2\pi R^2}$	$1+\frac{6(\lambda-p)}{5\pi R^2}$	$1+\frac{\lambda-p}{\pi R^2}$

* Neyman's Type A distribution. † Neyman's Type C distribution.

Except between the last three columns there is no strict basis of comparison between the indices of dispersion as they stand, since the first two functions permit theoretically infinite spreading. Of course comparisons within any column can be made, from which it follows that the relative degrees of contagion for the final distributions are in the same order as for the $p(n)$, the indices of dispersion for which are respectively $1+\lambda$, $1+\frac{1}{2}\lambda$, 1, $1-p$. With the distributions in the last three columns there is a definite 'clump' size with area equal to πR^2, the size of the circle outside which no members of a group fall. For the first two we can only specify some area containing a given percentage of the distribution as the clump size, and in the table below the indices of dispersion have been made comparable by taking as the clump size the area A containing on the average 95 % of the distribution. The values of A for the five $P(r)$ are approximately $5{\cdot}99\pi\sigma^2$, $22{\cdot}50\pi\sigma^2$, $0{\cdot}75\pi R^2$, $0{\cdot}81\pi R^2$ and $0{\cdot}95\pi R^2$.

A reasonable conclusion to be drawn from this modified table is that for various types of spreading mechanisms the resulting distributions all have approximately the same degree of contagion, or in other words, there will not be much difference in the forms of the distributions and the differences might even

be undetectable from samples. The only possible exception occurs in column (2), and this may be due to the exponential form of $P(r)$, with its slow falling off in frequency and consequent difficulty in defining any sort of clump size.

Table 2. *Modified values of the index of dispersion* μ_2/μ_1'

$p(n)$ \ $P(r)$	(1)	(2)	(3)	(4)	(5)
(1)	$1+3{\cdot}00\lambda/A$	$1+5{\cdot}63\lambda/A$	$1+2{\cdot}24\lambda/A$	$1+1{\cdot}95\lambda/A$	$1+1{\cdot}90\lambda/A$
(2)	$1+2{\cdot}25\lambda/A$	$1+4{\cdot}22\lambda/A$	$1+1{\cdot}68\lambda/A$	$1+1{\cdot}47\lambda/A$	$1+1{\cdot}43\lambda/A$
(3)	$1+1{\cdot}50\lambda/A$	$1+2{\cdot}81\lambda/A$	$1+1{\cdot}12\lambda/A$	$1+0{\cdot}98\lambda/A$	$1+0{\cdot}95\lambda/A$
(4)	$1+1{\cdot}50(\lambda-p)/A$	$1+2{\cdot}81(\lambda-p)/A$	$1+1{\cdot}12(\lambda-p)/A$	$1+0{\cdot}98(\lambda-p)/A$	$1+0{\cdot}95(\lambda-p)/A$

We may note in connexion with similarity of distributions that this is further borne out by a study of higher moments. For example, with the distributions in row (1), column (3) and row (2), column (1), which have almost identical μ_2/μ_1', we have the following values for μ_3/μ_1' and μ_4/μ_1':

$$[(1),(3)]\begin{cases}\mu_3/\mu_1' = 1+6{\cdot}73\lambda/A+9{\cdot}06(\lambda/A)^2,\\ \mu_4/\mu_1' = 1+15{\cdot}70\lambda/A+54{\cdot}34(\lambda/A)^2+54{\cdot}17(\lambda/A)^3;\end{cases}$$

$$[(2),(1)]\begin{cases}\mu_3/\mu_1' = 1+6{\cdot}74\lambda/A+8{\cdot}97(\lambda/A)^2,\\ \mu_4/\mu_1' = 1+15{\cdot}73\lambda/A+53{\cdot}85(\lambda/A)^2+50{\cdot}41(\lambda/A)^3.\end{cases}$$

5. Thomas's double Poisson distribution has mean $m(1+\lambda)$ and variance $m(1+3\lambda+\lambda^2)$ for the distribution of parents and offspring, so that for the offspring only $\mu_1' = m\lambda$, $\mu_2/\mu_1' = 3+\lambda$. In general therefore the degree of contagion will tend to be higher than for the distributions above if A is more than 1 quadrat. This arises because Thomas assumes that all offspring of a group fall in the same quadrat as their parent, which is obviously an unrealistic assumption whatever the dispersal mechanism because of edge effects between neighbouring quadrats. It is possible with the radial Gaussian model to calculate the theoretical correlations between the numbers of offspring in neighbouring quadrats. (The basic theory enabling this calculation will be presented more fully in a later paper which will consider the general relations between plants in the quadrats of a grid superimposed upon the sampling region.) This is done in Table 3 for the correlations ρ_{01} between adjacent quadrats in the same row (or equivalently, column), and ρ_{11} between diagonally adjacent quadrats, with $\lambda = 3$ (the value of m is immaterial) and different values of h^2/σ^2. ($p(n)$ is assumed to follow the Poisson distribution.)

Table 3

h/σ	1	2	4	8
A	$18{\cdot}8h^2$	$4{\cdot}7h^2$	$1{\cdot}2h^2$	$0{\cdot}3h^2$
ρ_{01}	0·143	0·198	0·119	0·057
ρ_{11}	0·114	0·095	0·023	0·005

Thus there is still an appreciable boundary effect due to averaging even if most of the group is expected to fall within one quadrat; so that even if Thomas's distribution does give adequate fits to data, it is at best only an approximation to the facts and cannot be regarded as representing a very precise model for the spread of plants.

REFERENCES

Darwin, J. H. (1951). Unpublished Ph.D. Thesis, Manchester University.

Neyman, J. (1939). On a new class of 'contagious' distributions applicable in entomology and bacteriology. *Ann. Math. Statist.* **10**, 35.

Skellam, J. G. (1952). Studies in statistical ecology. I. Spatial pattern. *Biometrika*, **39**, 346.

Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika*, **36**, 18.

A general expression for the mean in a simple stochastic epidemic

BY H. W. HASKEY, *South-east Essex Technical College*

1. A community contains n susceptibles into which a single infectious individual is introduced and a mild infection spreads by contact between its members. It is supposed that no cases are withdrawn by isolation or death and that no case becomes clear of infection during the main part of the epidemic.

If $p_r(t)$ is the probability that there are r susceptibles still uninfected at time t, Bailey (1950) gives the following equations for the epidemic process:

$$\left.\begin{aligned} \frac{dp_r(t)}{dt} &= (\mathbf{r+1})p_{r+1}(t) - \mathbf{r}p_r(t) \quad (r = 0, 1, 2, \ldots (n-1)), \\ \text{and} \qquad \frac{dp_n(t)}{dt} &= -\mathbf{n}p_n(t), \end{aligned}\right\} \tag{1}$$

where $\mathbf{r} = r(n-r+1) = \mathbf{n-r+1}$. He shows that the rate of change of the mean number of infected at time t in this stochastic process is

$$\sum_{\mathbf{r}} \{\mathbf{r}f'_r(1)\,t + \mathbf{r}g'_r(1) - f'_r(1)\}\, e^{-\mathbf{r}t}, \tag{2}$$

where $f'_r(x)$, $g'_r(x)$ are the derivatives with respect to x of polynomials $f_r(x)$, $g_r(x)$ satisfying the equations

$$x(1-x)f''_r(x) - n(1-x)f'_r(x) - \mathbf{r}f_r(x) = 0, \tag{3}$$

$$x(1-x)g''_r(x) - n(1-x)g'_r(x) - \mathbf{r}g_r(x) = -f_r(x). \tag{4}$$

Bailey obtained an expression for $g'_1(1)$ but not for $g'_r(1)$, and was unable to obtain a convenient one for $g_r(x)$. If a series solution of (4) is attempted for $n = 30$ the coefficients of terms in the fifteen series required are often of the order of 10^{14}, and as the value of one coefficient is used to determine the next, it is impracticable to evaluate $g'_r(1)\,(r = 1, 2, \ldots, 15)$ accurately enough with a machine of ordinary capacity.

The object of this paper is to obtain a general formula for $g'_r(1)$ for n even or odd, by means of which the rates of change of the mean number of infected for $n = 30$ at various times are computed more accurately and readily.

2. Bailey uses the Laplace transform and its inverse with respect to time given by

$$\left.\begin{aligned} q_r(\lambda) &= \int_0^\infty e^{-\lambda t} p_r(t)\, dt \quad (R(\lambda) > 0), \\ 2\pi i p_r(t) &= \int_{c-i\infty}^{c+i\infty} e^{\lambda t} q_r(\lambda)\, d\lambda, \end{aligned}\right\} \tag{5}$$

where c is positive and greater than the abscissae of all the residues. The transforms of the equations (1) are taken, using the boundary conditions $p_r(0) = 1$ if $r = n$ and $p_r(0) = 0$ if $r \neq n$, leading to recurrence relations between the q's and thence to the values, for even n

$$q_s = \frac{n!\,(n-s)!}{s!\,(\lambda+\mathbf{1})(\lambda+\mathbf{2})\ldots(\lambda+\mathbf{s-1})\{(\lambda+\mathbf{s})(\lambda+\mathbf{s+1})\ldots(\lambda+\tfrac{1}{2}\mathbf{n})\}^2}, \quad (\tfrac{1}{2}n \geqslant s > 0), \tag{6}$$

$$q_s = \frac{n!\,(n-s)!}{s!\,(\lambda+\mathbf{1})(\lambda+\mathbf{2})\ldots(\lambda+\mathbf{n-s+1})} \quad (s > \tfrac{1}{2}n). \tag{7}$$

The transform of the probability generating function

$$\Pi(x, t) = \sum_{s=0}^{n} x^s p_s(t) \tag{8}$$

is

$$\Pi^*(x, \lambda) = \sum_{s=0}^{n} x^s q_s(\lambda),$$

which by (6) and (7), can be written, for n even,

$$\Pi^*(x, \lambda) = \lambda^{-1} + \sum_{r=1}^{\frac{1}{2}n} \left\{ \frac{f_r(x)}{(\lambda+\mathbf{r})^2} + \frac{g_r(x)}{(\lambda+\mathbf{r})} \right\}, \tag{9}$$

where $f_r(x)$, $g_r(x)$ are the sums of coefficients of $1/(\lambda+\mathbf{r})^2$, $1/(\lambda+\mathbf{r})$ respectively in partial fraction expansions of $\sum_s x^s q_s(\lambda)$. Bailey has shown that $f_r(x)$, $g_r(x)$ satisfy (3) and (4) and that

$$f_r'(1) = \frac{n!\,(n-2r+1)^2}{(n-r)!\,(r-1)!}, \quad n \text{ even or odd.}$$

Still taking n even, we write

$$q_s = \sum_{k=1}^{\frac{1}{2}n} \frac{A_{sk}}{(\lambda+\mathbf{k})} + \sum_{k=s}^{\frac{1}{2}n} \frac{B_{sk}}{(\lambda+\mathbf{k})^2} \quad (\tfrac{1}{2}n \geqslant s), \tag{10}$$

$$q_s = \sum_{k=1}^{n-s+1} \frac{A_{sk}}{(\lambda+\mathbf{k})} \quad (s > \tfrac{1}{2}n), \tag{11}$$

and then

$$g_r(x) = \sum_{s=1}^{n-r+1} A_{sr} x^s,$$

giving

$$g_r'(1) = \sum_{s=1}^{n-r+1} s A_{sr}. \tag{12}$$

We proceed to find the A_{sr} and use (12).

3. Multiply (6) and (10) by $(\lambda+\mathbf{r})$ and then put $\lambda = -\mathbf{r}$. On simplification it is found that

$$A_{sr} = \frac{(-)^{r-1} n!\,(n-s)!\,(s-r-1)!\,(n-2r+1)}{s!\,(r-1)!\,(n-r)!\,(n \div 1-r-s)!} \quad (\tfrac{1}{2}n \geqslant s > r). \tag{13}$$

Treating (7) and (11) similarly the same result, (13), is obtained for $n-r+1 \geqslant s > \frac{1}{2}n$. From (7), $A_{sr} = 0$ if $s > n-r+1$. When $\frac{1}{2}n \geqslant s$ with $r \geqslant s$, there are squared factors in the denominator of (6), and it is seen that

$$A_{sr} = \frac{\partial}{\partial\lambda}\{q_s(\lambda+\mathbf{r})^2\}_{\lambda=-\mathbf{r}}.$$

Noting that

$$\frac{1}{\mathbf{s}-\mathbf{r}} = \frac{1}{(n-2r+1)}\left\{\frac{1}{(s-r)} + \frac{1}{(n+1-r-s)}\right\}, \tag{14}$$

it is found that

$$A_{sr} = \frac{(-)^s n!\,(n-s)!\,(n-2r+1)\left\{\sum_{p=r}^{n-r} p^{-1} + \sum_{p=r-s+1}^{n-r+s+1} p^{-1} - 2/(n-2r+1)\right\}}{s!\,(r-1)!\,(n-r)!\,(n+1-r-s)!\,(r-s)!}. \tag{15}$$

In (15) the first sum does not arise when $s = 1$ and the second does not arise if $s = \frac{1}{2}n$.

Equations (13) and (15) give, for even n, all types of term on the right of (12).

4. From (13) the contribution to $g_r'(1)$ due to terms for which $s > r$ is

$$\mu = (-)^{r-1} \frac{n!\,(n-2r+1)}{(n-r)!\,(r-1)!} \sum_{y=1}^{n-2r+1} \frac{(n-r-y)!\,(y-1)!}{(y+r-1)!\,(n+1-2r-y)!}, \tag{16}$$

where $s-r = y$. The general term in (16) can be put in partial fractions so that

$$\frac{(-)^{r-1}(n-r-y)!\,(y-1)!}{(y+r-1)!\,(n+1-2r-y)!} = \sum_{k=0}^{r-1} \frac{c_k}{y+k}, \tag{17}$$

where

$$c_k = \frac{(-)^{r+k-1}(n-r+k)!}{(n+1-2r+k)!\,(r-k-1)!\,k!}. \tag{18}$$

Writing the terms of $\sum_{k=0}^{r-1} c_k$ in the reverse order it is seen that

$$\begin{aligned}\frac{(n-r)!\,(r-1)!}{(n-1)!} \sum_{k=0}^{r-1} c_k &= 1 - \frac{(n-r)^{(1)}(r-1)^{(1)}}{(n-1)^{(1)}\,1!} + \frac{(n-r)^{(2)}(r-1)^{(2)}}{(n-1)^{(2)}\,2!} - \dots \\ &= F\{-(r-1), -(n-r);\ -(n-1), 1\} \\ &= \frac{(n-r)!\,(r-1)!}{(n-1)!};\end{aligned}$$

Hence

$$\sum_{k=0}^{r-1} c_k = 1. \tag{19}$$

5. The other contribution to $g'_r(1)$ due to terms for which $s \leqslant r$ is $\sum_{s=1}^{r} sA_{sr}$, where sA_{sr}, from (15) can, from (18), be written

$$uvc_{r-s} + uc_{r-s} \sum_{p=r-s+1}^{n-r-s+1} p^{-1},$$

where $$v = \sum_{p=r}^{n-r} p^{-1} - 2/(n-2r+1), \quad u = -n!\,(n-2r+1)/(n-r)!\,(r-1)!.$$

Writing out the values of this for $s = 1, 2, \ldots, r$ on separate lines as follows

$$uvc_{r-1} + uc_{r-1}[1/r + 1/(r+1) + \ldots + 1/(n-r)],$$
$$uvc_{r-2} + uc_{r-2}[1/(r-1) + 1/r + \ldots + 1/(n-r-1)],$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$uvc_0 + uc_0[1/1 + \tfrac{1}{2} + \ldots + 1/(n-2r+1)],$$

and summing by columns there results

$$uv \sum_{k=0}^{r-1} c_k + u \sum_{k=0}^{r-1} \left(\frac{c_k}{1+k} + \frac{c_k}{2+k} + \ldots + \frac{c_k}{n-2r+1+k} \right),$$

or by (19),
$$uv + u \sum_{y=1}^{n-2r+1} \sum_{k=0}^{r-1} c_k/(y+k) \tag{20}$$

or
$$uv - \mu \tag{21}$$

by (16) and (17). Then $g'_r(1)$, sum of (16) and (20), is by (21)

$$g'_r(1) = \frac{n!}{(n-r)!\,(r-1)!} \left\{ 2 - (n-2r+1) \sum_{p=r}^{n-r} \frac{1}{p} \right\} \quad (r = 1, 2, \ldots, \tfrac{1}{2}n). \tag{22}$$

With n odd the foregoing method applies but there are modifications, e.g.

$$q_s = \frac{n!\,(n-s)!}{s!\,(\lambda+1)\,(\lambda+2)\ldots(\lambda+s-1)\left\{(\lambda+s)\,(\lambda+s+1)\ldots\left(\lambda+\frac{n-1}{2}\right)\right\}^2\left(\lambda+\frac{n+1}{2}\right)} \quad \left(\frac{n+1}{2} > s\right),$$

and (15) is replaced by

$$A_{sr} = \frac{(-)^s n!\,(n-s)!\,(n-2r+1)\left\{\sum_{k=1}^{s-1} \frac{1}{v} + 2 \sum_{k=s,\,k\neq r}^{\frac{1}{2}(n-1)} \frac{1}{v} + \frac{4}{(n-2r+1)}\right\}}{s!\,(r-1)!\,(n-r)!\,(r-s)!\,(n+1-r-s)!}$$

for $\frac{1}{2}(n-1) \geqslant r \geqslant s$, where $1/v = 1/(k-r) + 1/(n+1-k-r)$. Equation (13) applies if $s > r$ and, when $r = \frac{1}{2}(n+1)$ a summation analagous to that by which $\sum_{k=0}^{r-1} c_k$ was obtained is needed. When $s \geqslant \frac{1}{2}(n+1)$, (7) holds so that $f'_{\frac{1}{2}(n+1)}(1)$ is zero. It is found that, for n odd,

$$g'_r(1) = \frac{n!}{(r-1)!\,(n-r)!} \left\{ 2 - (n-2r+1) \sum_{p=r}^{n-r} \frac{1}{p} \right\} \quad (r = 1, 2, \ldots, \tfrac{1}{2}(n-1)), \tag{23}$$

but
$$g'_{\frac{1}{2}(n+1)}(1) = \frac{n!}{\{[\frac{1}{2}(n-1)]!\}^2},$$

which is only half that given by the general formulae (22) and (23).
Thus the Stochastic mean number uninfected at time t is

$$\sum_r \frac{n!}{(n-r)!\,(r-1)!} \left\{ (n-2r+1)^2 t + 2 - (n-2r+1) \sum_{p=r}^{n-r} \frac{1}{p} \right\} e^{-rt}.$$

In this formula $r = 1, 2, \ldots, \frac{1}{2}n$ if n is even, but if n is odd $r = 1, 2, \ldots, \frac{1}{2}(n+1)$ with the special end-value given above.

6. The transform of (9) gives

$$\Pi(x, t) = 1 + \sum_{r=1}^{l} \{t f_r(x) + g_r(x)\} e^{-rt}, \tag{24}$$

where $l = \frac{1}{2}n$ is n if even and $l = \frac{1}{2}(n+1)$ if n is odd. Hence the rate of change of the mean number of infected at time t is given by $-\left.\dfrac{\partial^2 \Pi}{\partial t\,\partial x}\right|_{x=1}$, the value of which is given by (2). For $n = 30$, the values of $\mathbf{r}f'_r(1)$ accurately and $-\mathbf{r}g'_r(1) + f'_r(1)$ to three decimal places are given in Table 1.

Table 1

r	$rf'_r(1)$	$\{-rg'_r(1)+f'_r(1)\}\times 10^3$
1	756900	126829916
2	36785340	4521346367
3	639450000	66734940835
4	6262809840	584644959146
5	40849344900	3499281732477
6	192917497500	15372766370623
7	691895131200	51600518480155
8	1938488760000	135195958672714
9	4308712679700	278181087951791
10	7634438311500	443723391803482
11	10708043346000	520480237463232
12	11595627716400	364795469521989
13	9107736592500	177034391329189
14	4360935177900	659179172373612
15	858423072000	1077291176400000

Chambers's shorter tables were used to give e^{-x^2} to six decimal places enabling z, the rate of change of the mean number of infected, to be found to the nearest integer for times $\geqslant 0{\cdot}08$.

Table 2 *Stochastic treatment*, $n = 30$

Time	0	0·05	0·06	0·08	0·1	0·11	0·12	0·15	0·2	0·3	0·4
z	30	104	118	152	176	183	183	150	90	12	1

The deterministic expression for z is

$$n(n+1)^2 e^{(n+1)t}/\{n+e^{(n+1)t}\}^2,$$

giving a maximum of $\frac{1}{4}(n+1)^2$ at time $\log_e n/(n+1)$.

Table 3. *Deterministic treatment*, $n = 30$

Time	0·005	0·01	0·02	0·03	0·05	0·06	0·07	0·09	0·1	0·12	0·13	0·15	0·18	0·2	0·25	0·28	0·1097
z	35	40	53	69	113	140	168	219	235	234	218	166	94	52	12	4	240¼

The deterministic and stochastic curves for z ($n = 30$) show the same characteristics as those for $n = 10$, 20; their maxima occur at about the same time, less than that for $n = 20$, and the difference between the maxima exceeds that for $n = 20$.

REFERENCE

BAILEY, N. T. J. (1950). *Biometrika*, **37**, 193.

Some remarks on confidence or fiducial limits

BY THEODORE E. STERNE

Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland

If the probability of a 'success' is π, then the probability that there will be precisely r successes during n trials is exactly

$$p_{n,r}(\pi) = \frac{n!}{r!\,(n-r)!}\,\pi^r(1-\pi)^{n-r}. \qquad (1)$$

Therefore, the probability of obtaining a number of successes as probable as or less probable than a particular number of successes s is equal to

$$P(\pi, n, s) = \sum_r p_{n,r}(\pi), \qquad (2)$$

where the summation is over all values of r for which

$$p_{n,r}(\pi) \leqslant p_{n,s}(\pi). \tag{3}$$

The probability of obtaining a value of P equal to or less than some particular and possible value, P', of P is equal to P'. From the preceding considerations, confidence or fiducial limits of π can be chosen corresponding to any values of n and s and to any desired value, ϵ lying between 0 and 1, of P.

Consider the dependence of $P(\pi, n, s)$ upon the variable π, for any value of s other than 0 or n. For values of s differing from 0 and n it is clear that $P(0, n, s)$ and $P(1, n, s)$ are zero, and that there are some values of π, between 0 and 1, for which $P(\pi, n, s)$ is unity. The values of π for which $P(\pi, n, s)$ is unity are values for which $p_{n,s}(\pi)$ is the largest of all terms $p_{n,r}(\pi)$. It can readily be shown that there are exactly n values of π between 0 and 1 at which $P(\pi, n, s)$ is discontinuous. Such discontinuities correspond to the 'crossing' of $p_{n,s}(\pi)$ by other $p_{n,r}(\pi)$'s, and except at the n discontinuities $P(\pi, n, s)$ is a continuous function of π. When s is 0, the situation is somewhat different since then $P(0, n, 0)$ is unity while $P(1, n, 0)$ is zero. When s equals n, $P(0, n, n)$ is zero while $P(1, n, n)$ is unity. Whatever may be the value of s, there are just n values of π at which $P(\pi, n, s)$ is discontinuous.

The preceding description of some of the properties of the P-function has been included to clarify, in the reader's mind, the nature of the dependence of $P(\pi, n, s)$ upon π. To select confidence or fiducial limits,

$$p_1 \leqslant \pi \leqslant p_2,$$

it is sufficient to consider the shortest interval of π that contains all the values of π for which

$$P(\pi, n, s) \geqslant \epsilon, \tag{4}$$

corresponding to any particular and possible values of n and s, and to some desired value of ϵ. Corresponding to any possible values n, s and ϵ it is always possible to find the shortest interval, (p_1, p_2), by calculation. The lower limit p_1 is the smallest of all π's satisfying the relation (4), and the upper limit p_2 is the largest of all π's satisfying the relation (4). Such limits p_1 and p_2 always exist, as will be shown. One may note first that when s is zero, p_1 is zero; and that when s is n, p_2 is unity.

The existence of p_1 and p_2 follows rigorously from the consideration that the set of π's satisfying (4) has a lower bound, zero, and must therefore have a greatest lower bound, p_1. Similarly, the set of π's satisfying (4) has an upper bound, unity, and must therefore have a least upper bound, p_2. It can further be shown that because of the nature of the P-function and of its discontinuities, the numbers p_1 and p_2 must themselves satisfy (4). Because of the last-mentioned property of p_1 and p_2, it follows that

$$P(\pi, n, s) < \epsilon \tag{5}$$

for all values of π lying outside the closed interval $p_1 \leqslant \pi \leqslant p_2$. Let us now arbitrarily choose some value of ϵ between 0 and 1 and adopt the policy, whenever in an experiment some value of s has been observed, of always asserting that

$$p_1 \leqslant \pi \leqslant p_2, \tag{6}$$

where p_1 and p_2 correspond to the actual n, the observed s, and the chosen ϵ. We shall be wrong in our assertions when, and only when, π lies outside the closed interval (6), in which case $P < \epsilon$ by (5). The probability that our assertion is incorrect, which is the same as the proportion of times that we shall be wrong in such assertions in the long run, is therefore less than ϵ. Therefore the probability that our assertion (6) is correct, which is the same as the proportion of times that we shall in the long run be right in such assertions, is greater than $1 - \epsilon$. We accordingly call p_1 and p_2 the lower and upper $1 - \epsilon$ confidence, or fiducial, limits.

In the following tables are listed confidence limits p_1 and p_2 for values of ϵ equal to 0·5 and 0·10. These limits may be called the 50 and 90 % confidence limits. The tables cover all possible values of s corresponding to values of n from 1 through 10.

The author is inclined to prefer confidence limits, selected in accordance with the procedure given in this paper to confidence limits selected in accordance with the procedure of Clopper and Pearson (1934). Our discussion has considered the probability $P(\pi, n, s)$ of obtaining a number of successes as probable as or less probable than an observed number of successes s. Their discussion considers the probability of obtaining a number of successes equal to or less than the observed number and, separately, the probability of obtaining a number of successes equal to or greater than the observed number. They select an upper confidence limit in such a way as to make the first probability less than or equal to $\frac{1}{2}\epsilon$, and a lower limit in such a way as to make the second probability also less than or equal to $\frac{1}{2}\epsilon$. There is, however, no value of π that can make the second probability less than or equal to $\frac{1}{2}\epsilon$ when s is zero, nor is there any value of π that can make the first probability less than or equal to $\frac{1}{2}\epsilon$ when s is n. Consequently, the confidence limits selected by Clopper & Pearson are unnecessarily wide when s is either 0 or n, and they have in effect over-modestly underestimated the legitimate degree of confidence in their assertions,

comparable to (6), for values of s equal to 0 and n. The confidence intervals defined here are certainly narrower than theirs when s is 0 or n, and appear to be narrower in other cases when n is small. The advantage of the present limits over those of Clopper & Pearson is probably small when n is large. The present intervals are not 'central', in the sense in which the intervals of Clopper & Pearson are central, but they are perhaps more nearly 'central' than those of Clopper & Pearson with respect to the probabilities, $P(\pi, n, s)$, of obtaining results s as probable as or less probable than some observed s, regardless of whether such less probable s's are larger or smaller than the observed s.

Summary. In a binomial problem, confidence limits of an unknown parent probability can be selected on the basis of the probability, P, of obtaining a number of successes as probable as, or less probable than, the observed number s of successes in n trials. In comparison with confidence intervals

Table giving confidence limits for π for samples of size n

s	50%		90%	
	p_1	p_2	p_1	p_2
		$n = 1$		
0	0·00	0·50	0·00	0·90
1	·50	1·00	·10	1·00
		$n = 2$		
0	0·00	0·50	0·000	0·68
1	·29	·71	·051	·949
2	·50	1·00	·32	1·00
		$n = 3$		
0	0·00	0·37	0·000	0·54
1	·21	·63	·035	·80
2	·37	·79	·20	·965
3	·63	1·00	·46	1·00
		$n = 4$		
0	0·00	0·29	0·000	0·50
1	·16	·50	·026	·68
2	·29	·71	·14	·86
3	·50	·84	·32	·974
4	·71	1·00	·50	1·00
		$n = 5$		
0	0·00	0·24	0·000	0·40
1	·13	·41	·021	·60
2	·24	·59	·11	·75
3	·41	·76	·25	·89
4	·59	·87	·40	·979
5	·76	1·00	·60	1·00
		$n = 6$		
0	0·00	0·21	0·000	0·34
1	·11	·35	·017	·54
2	·21	·50	·093	·67
3	·35	·65	·20	·80
4	·50	·79	·33	·907
5	·65	·89	·46	·983
6	·79	1·00	·66	1·00

s	50%		90%	
	p_1	p_2	p_1	p_2
		$n = 7$		
0	0·000	0·18	0·000	0·35
1	·094	·31	·015	·50
2	·18	·44	·079	·65
3	·31	·56	·17	·72
4	·44	·69	·28	·83
5	·56	·82	·35	·921
6	·69	·906	·50	·985
7	·82	1·00	·65	1·00
		$n = 8$		
0	0·000	0·16	0·000	0·31
1	·083	·27	·013	·44
2	·16	·39	·069	·56
3	·27	·51	·15	·69
4	·39	·61	·24	·76
5	·49	·73	·31	·85
6	·61	·84	·44	·931
7	·73	·917	·56	·987
8	·84	1·00	·69	1·00
		$n = 9$		
0	0·000	0·14	0·000	0·28
1	·074	·25	·012	·39
2	·14	·37	·061	·52
3	·25	·50	·13	·61
4	·34	·59	·21	·72
5	·41	·66	·28	·79
6	·50	·75	·39	·87
7	·63	·86	·48	·939
8	·75	·926	·61	·988
9	·86	1·00	·72	1·00
		$n = 10$		
0	0·000	0·13	0·000	0·25
1	·067	·22	·010	·40
2	·13	·36	·055	·50
3	·22	·45	·12	·60
4	·30	·55	·19	·66
5	·36	·64	·25	·75
6	·45	·70	·34	·81
7	·55	·78	·40	·88
8	·64	·87	·50	·945
9	·78	·933	·60	·990
10	·87	1·00	·75	1·00

selected in accordance with procedures of Clopper & Pearson, the present intervals appear to be sometimes narrower. Tables are given of the 50 and 90 % confidence limits of the parent probability for values of n up to 10.

REFERENCE

CLOPPER, C. J. & PEARSON, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, **26**, 404–13.

A note on the probability integral of the correlation coefficient

BY B. I. HARLEY, *University College, London*

1. Tables of the probability integral of r, the correlation coefficient, have been given by David (1938) for the population correlation coefficient $\rho = 0{\cdot}0\,(0{\cdot}1)\,0{\cdot}9$ and the sample size $n = 3\,(1)\,25,\ 50,\ 100,\ 200,\ 400$. In the introduction to the tables it was shown that the normalizing transformation, $z = \tanh^{-1} r$, given by R. A. Fisher (1921), provided adequate approximation to the integral for a large range of values of ρ, but that for ρ in the neighbourhood of 0·9 the results became inaccurate. Recently values of the integral were required for $\rho = 0{\cdot}9$ and n between 25 and 50, i.e. beyond the tabular range, and it was considered that the z-transformation would not give sufficient accuracy. An examination of the recurrence formulae suggested by F. Garwood (1933) and recently by H. Hotelling (1953) did not promise any greater accuracy, and some fresh method of approximation was looked for.

2. Since the z-transformation is a normalizing one, the distribution of

$$z = \tfrac{1}{2}\log_e \frac{1+r}{1-r}$$

may be expected to be more nearly normal than that of r, whatever the value of ρ. It appears reasonable, therefore, to use the cumulants of z in an expansion such as that given by Edgeworth, since this expansion is known to give accurate results in the neighbourhood of the normal. The first four cumulants of z in samples of n from a normal bivariate population are found from the values of the moments of z given by A. K. Gayen (1951) to be:

$$\kappa_1(z) = \tfrac{1}{2}\log_e \frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)}\left\{1 + \frac{5+\rho^2}{4(n-1)} + \ldots\right\}, \tag{1}$$

$$\kappa_2(z) = \frac{1}{n-1}\left\{1 + \frac{4-\rho^2}{2(n-1)} + \frac{22-6\rho^2-3\rho^4}{6(n-1)^2} + \ldots\right\}, \tag{2}$$

$$\kappa_3(z) = \frac{\rho^3}{(n-1)^3} + \ldots, \tag{3}$$

$$\kappa_4(z) = \frac{1}{(n-1)^3}\left\{2 + \frac{3(4-\rho^4)}{n-1} + \ldots\right\}. \tag{4}$$

Substituting these values in the Cornish-Fisher (1937) form of the Edgeworth expansion, taking a preliminary mean and variance as

$$m = \tfrac{1}{2}\log_e \frac{1+\rho}{1-\rho}, \quad v = (n-1)^{-1},$$

we obtain the probability integral of z

$$\begin{aligned}\int_{-\infty}^{Z} z\,dz = {} & \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{x} e^{-\frac{1}{2}x^2}\,dx + \frac{1}{\sqrt{(2\pi)}}\,e^{-\frac{1}{2}x^2}\left\{\frac{m-\kappa_1}{v^{\frac{1}{2}}} + \frac{H_1}{2v}[(\kappa_1-m)^2+(\kappa_2-v)]\right.\\
& - \frac{H_2}{6v^{\frac{3}{2}}}[\kappa_3+(\kappa_1-m)^3+3(\kappa_1-m)(\kappa_2-v)]\\
& + \frac{H_3}{24v^2}[4\kappa_3(\kappa_1-m)+\kappa_4+(\kappa_1-m)^4+3(\kappa_2-v)^2+6(\kappa_1-m)^2(\kappa_2-v)] - \frac{H_4}{24v^{\frac{3}{2}}}[\kappa_4(\kappa_1-m)]\\
& \left. + \frac{H_5}{48v^3}[(\kappa_1-m)^2\kappa_4+(\kappa_2-v)\kappa_4] + \frac{H_7}{1152v^4}\kappa_4^2 + \ldots\right\},\end{aligned} \tag{5}$$

where

$$Z = \tfrac{1}{2}\log_e \frac{1+R}{1-R} = m + xv^{\frac{1}{2}},$$

and R is the particular value of r for which the probability integral is required.

$\kappa_1, \kappa_2, \kappa_3, \kappa_4$ are the first four cumulants of z which are tabulated above in equations (1)–(4). H_i, $i = 1, 2, \ldots, 7$ are the Hermite polynomials given by

$$\frac{d^i}{dx^i}e^{-\frac{1}{2}x^2} = H_i e^{-\frac{1}{2}x^2},$$

$$H_1 = -x, \qquad H_4 = x^4 - 6x^2 + 3,$$
$$H_2 = x^2 - 1, \qquad H_5 = -x^5 + 10x^3 - 15x,$$
$$H_3 = -x^3 + 3x, \qquad H_7 = -x^7 + 21x^5 - 105x^3 + 105x.$$

The expansion includes terms up to order n^{-2}.

3. For the purpose of testing the accuracy of the formula we took $\rho = 0{\cdot}9$ and $n = 25$ and 50. The value of the probability integral of r, obtained by using the normal z-transformation and the values of $\kappa_1(z)$ and $\kappa_3(z)$ given in equations (1) and (2), was also calculated. Some results of calculations are given in Table 1.

Table 1. *Comparison of exact and approximate values of the probability integral of r*

(i) $n = 25$, $\rho = 0{\cdot}9$.

r	Exact value	Approximate value from equation (5)	Approximate value from z-transformation
0·75	0·00 743	0·00 742	0·00 698
0·82	0·05 574	0·05 578	0·05 589
0·84	0·09 859	0·09 862	0·09 981
0·87	0·22 387	0·22 379	0·22 576
0·90	0·46 244	0·46 247	0·46 250
0·93	0·78 645	0·78 661	0·78 442
0·95	0·94 612	0·94 612	0·94 609
0·965	0·99 263	0·99 264	0·99 325

(ii) $n = 50$, $\rho = 0{\cdot}9$.

r	Exact value	Approximate value from equation (5)	Approximate value from z-transformation
0·82	0·01 285	0·01 286	0·01 265
0·85	0·05 998	0·06 001	0·06 024
0·88	0·23 202	0·23 200	0·23 294
0·90	0·47 403	0·47 404	0·47 405
0·92	0·77 108	0·77 112	0·77 009
0·93	0·88 871	0·88 872	0·88 814
0·95	0·99 174	0·99 174	0·99 204

It will be noted that the agreement between the value from equation (5) and the exact value from the tables is even better for $n = 50$ than for $n = 25$, and the results suggest that the formula can be used with confidence for n between 25 and 50.

4. The inverse Cornish-Fisher expansion can be used to obtain the significance levels for any n. Inverting (5) we have for standardized z, say z',

$$z' - \xi_\alpha = \frac{\kappa_1 - m}{v^{\frac{1}{2}}} + \frac{(\kappa_2 - v)}{2v}\xi_\alpha + \frac{\kappa_4}{24v^2}(\xi_\alpha^3 - 3\xi_\alpha) + \frac{\kappa_3}{6v^{\frac{3}{2}}}(\xi_\alpha^2 - 1)$$
$$- \frac{(\kappa_2 - v)^2}{8v^2}\xi_\alpha - \frac{(\kappa_2 - v)}{16v^2}\kappa_4(\xi_\alpha^3 - 3\xi_\alpha) - \frac{\kappa_4^2}{384v^4}(32\xi_\alpha^5 - 24\xi_\alpha^3 + 29\xi_\alpha),$$

where ξ_α is the unit normal deviate cutting off $100\alpha\%$ from the upper tail. $\kappa_1, \kappa_2, \kappa_3, \kappa_4$ are the first four cumulants of z. The terms are arranged in descending order in n, and only those up to order n^{-2} have been retained.

For $n = 25$, $\rho = 0{\cdot}9$ we obtain, by this means, a value of 1·83965 for the 5 % significance level of z, or 0·95076 for r, a value whose accuracy could probably not be reached by interpolation in David's tables.

For $n = 50$, $\rho = 0{\cdot}9$ the 5 % significance level is 1·72055 for z, and 0·93793 for r.

REFERENCES

Cornish, E. A. & Fisher, R. A. (1937). Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Statist.* **4**, 307.

David, F. N. (1938). *Tables of the Ordinates and Probability Integral of the distribution of the Correlation Coefficient in small samples.* Cambridge University Press.

Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, **1**, no. 4, 3.

Garwood, F. (1933). The probability integral of the correlation coefficient in samples from a normal bivariate population. *Biometrika*, **25**, 71.

Gayen, A. K. (1951). The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. *Biometrika*, **38**, 219.

Hotelling, H. (1953). New light on the correlation coefficient and its transforms. *J. R. Statist. Soc.* B, **15**, 193.

A simplified expression for the variance of the χ^2-function on a contingency table

By REED B. DAWSON, Jr., *U.S. Department of Defense*

Haldane (1940) considers the $(m \times n)$-fold contingency table

a_{11}	a_{12}	...	a_{1m}	s_1
a_{21}	a_{22}	...	a_{2m}	s_2
.....	.....	...	.....	...
a_{n1}	a_{n2}	...	a_{nm}	s_n
t_1	t_2	...	t_m	N

and the usual χ^2-function in the form

$$\chi^2 = N\Sigma\frac{a_{ij}^2}{s_i t_j} - N.$$

Let

$$\nu = \frac{(n-1)(N-n)}{N-1}, \quad \mu = \frac{(m-1)(N-m)}{N-1},$$

$$\sigma = \frac{N\Sigma s_i^{-1} - n^2}{N-2}, \quad \tau = \frac{N\Sigma t_i^{-1} - m^2}{N-2}.$$

Then the variance may be found from the formula

$$V(\chi^2) = \frac{2N}{N-3}(\nu-\sigma)(\mu-\tau) + \frac{N^2}{N-1}\sigma\tau,$$

whose identity with the formula (4) given by Haldane may be verified directly. Note that whenever the row (or column) sums are equal σ (or τ) vanishes, eliminating the second term.

REFERENCE

Haldane, J. B. S. (1940). The mean and variance of χ^2, when used as a test of homogeneity, when expectations are small. *Biometrika*, **31**, 346–55.

REVIEWS

Elementary Medical Statistics. By D. MAINLAND. London: W. B. Saunders Company Ltd. 1952. Pp. 327. 25*s*.

The term 'statistics' is even less precisely defined than usual when prefaced by the epithet 'medical'. Thus, a 'medical statistician' may be concerned with the application of standard biometric techniques in clinical and laboratory research; he may be engaged in the study of large-scale mortality and morbidity statistics, where sampling errors are usually unimportant; or he may have the largely administrative task of organizing and working a hospital records system. This book deals mainly with the first field of activity, and will interest medical research workers who have felt the need for a non-mathematical book on elementary statistical methods with examples drawn from the medical sphere.

It is gratifying to find shrewd discussions on a number of problems which occur in medical statistics, and for which the research worker will find little guidance in the standard books. In clinical medicine, from which most of the examples are drawn, rigorous experimental methods are often, though not always, impossible, and Prof. Mainland carefully points out the dangers inherent in the interpretation of fortuitously collected data. He has some wise words, too, on the meaning (or lack of meaning) of so-called 'normal' values, and on errors in routine procedures like blood counting. His short sections on vital statistics, however, are less happy; the definitions of stillbirth and maternal mortality rates given on p. 36 do not apply in this country (where the denominator in each rate should be total births, live and still), and the expectation of life is wrongly interpreted on p. 255.

The treatment of standard statistical methodology, although sound, is unfortunately marred by a curiously disconnected exposition. A very good introductory section is followed by a chapter 'On looking at evidence', in which a bewildering number of hares are chased and hastily dropped. Then follow 100 pages nominally on enumeration data, but in which is included an excellent discussion of the role of randomization in experimental design. Also included here is an exposition of confidence limits and significance tests for the binomial distribution—in that order, surprisingly; the usual difficulty in understanding confidence estimation is likely to be increased when the methods are first illustrated for a discrete distribution like the binomial. Two chapters on measurements, including a detailed treatment of the *t*-test, are followed by a chapter on statistical ideas in clinical practice, some features of which have already been mentioned, and a final chapter briefly introducing one-way and two-way analysis of variance, regression and correlation, probit and other transformations, and many other topics. There is a great deal of repetition throughout the book. Hints are constantly dropped about matters which are to be discussed in more detail later, and the reader is frequently referred to forthcoming sections.

These faults make the book unsuitable as a primer in statistical methods, but it would be a pity if they prevented the medical research worker from reading the many excellent sections in it. He would benefit most from this book, if, before reading it, he first obtained a grounding in statistical methods from some other elementary text.

A feature of the book which should interest the general statistician is the extensive series of tables and charts for binomial confidence limits.

P. ARMITAGE

A Statistical Primer. By F. N. DAVID. London: Charles Griffin and Co. Ltd. 1953. Pp. x + 226. 22*s*.

This book is intended for the non-mathematical research worker who wishes to learn some statistics, enough to enable him to perform ordinary tests and to design satisfactory experiments without other expert help.

The non-mathematician is a timid creature, easily scared by the sudden appearance of an algebraic formula, unable to trust even the clearest mathematical demonstration; but quite able to 'get the idea' if it is suitably presented to him. It is fascinating to watch how carefully the author stalks her prey in the excellently written early chapters of this work, avoiding any sudden movement that might cause him to fly off in alarm. But alas, statistics is really quite a difficult subject, and soon the reader is grappling with a straightforward, and therefore rather difficul., chapter on moments.

The book then develops in a systematic way the various tests of significance—of a single mean and variance, then of paired means and variances. There is a wealth of numerical illustration to make

calculation more certain, and the author takes special pains over a number of points which often give worry to a beginner—for example, the question of one or two tails and of $(N-1)$ instead of N in a denominator (though the frightening expression 'degrees of freedom' is introduced suddenly and unceremoniously!).

Although the emphasis throughout is on elementary principles it is noticeable that there are no half-truths. Exposition is always rigorous with laudable emphasis on basic assumptions; for example, the effects of non-normality are described in connexion with each statistical test.

The final chapters deal with the analysis of variance (in a very clear and simple way), with the binomial and Poisson distributions and with the χ^2 test. Since all the other familiar tests have been covered it is somewhat surprising to find no mention of the correlation coefficient.

Short tables of the important statistical functions are included so this book can provide the complete armoury required for simple statistical work in the laboratory. D. R. WILKIE

Statistical Methods in Electrical Engineering. By D. A. BELL. London: Chapman and Hall. 1953. Pp. viii + 175. 25*s*.

It is refreshing to find the author of a book of this type remarking: 'Since one of the difficulties of the beginner in statistics is that most of the introductory books are written for specialist fields... it seems at first a retrograde step to write another book "for engineers".' Such an admission should be treasured by those who believe that many difficulties in statistics arise from premature concern with specific applications before basic theory has been thoroughly assimilated. In the present book, however, a good case is made out for specialization, on account of the particular *theoretical* developments of interest to the electrical engineer which find little or no place in elementary statistical theory. Thus the author devotes considerable attention to auto-correlation and power spectra, to fluctuations in electrical circuits, and to the use of indices of 'information' in communication engineering.

Unfortunately, the fact that concern with specialized theory is more than usually justifiable does not eliminate the difficulty of combining specialization with a clear development *ab initio* of fundamental statistical ideas. The small size of the book does not ease the task, but it does seem that good use has not been made of the available space. A general outline of the path followed by the author will elucidate this criticism. He starts with a discussion of probability theories, in the course of which the *frequency* theory is castigated for depending on the 'undefined concept' of 'equally probable'. There follows (Chapter II) a discussion of qualitative classification and, in particular, of contingency tables. The concept of significance is first mentioned in the following sentence: 'The idea of "degrees of freedom" is fundamental to the assessing of the "significance" of a variation, i.e. deciding whether an inference drawn from the statistics is likely to be valid or whether the observed variation is likely to be a random error of sampling.' An all too common beginner's error crops up on p. 22: '...if the observed value of χ^2, has, say, a 5% probability on the basis of a chance distribution there is a 95% probability that it is due to some association between the different attributes.'

Chapter III passes with bewildering brevity from observed distributions to binomial, Gaussian, Poisson, Rayleigh, Maxwell-Boltzmann and Fermi-Dirac distributions. Chapter IV proceeds at an equally breath-taking speed coveri..g moments, measures of dispersion, etc., and characteristic functions. In neither of these chapters is the distinction between sample and population made clear. Chapter V (Curve Fitting) discusses both the fitting of regression lines and distribution functions. Chapter VII (The Reliability of Data) is the only other chapter openly devoted to basic theory, though Chapter VIII (Principles of Quality Control) is in effect a belated discussion of elementary ideas.

It would be bad enough were the reader to be faced with the task of unscrambling correct ideas from confused presentation. There is, unfortunately, evidence that the author's own appreciation of statistical theory is at fault. One example has already been quoted. We further find that on p. 104 a t-statistic is referred to a t-distribution with 4 degrees of freedom because the arithmetic mean in the numerator is based on five observations, even though the standard deviation in the denominator is based on sixty observations. Confusion becomes worse confounded by a discussion which appears to mix up one-tail and two-tail tests. (Also the formula at the foot of p. 105 is in error by omission of $(2\pi)^{-\frac{1}{2}}$.) For a final example we look at pp. 101–2, where we find that the standard deviation of a mean of N (independent) observations is $1/\sqrt{(N-1)}$ times the standard deviation of each observation.

In view of features such as these the book cannot be recommended to engineers or others desiring a sound basic training in statistical theory. It may be of some value as a compendium of probabilistic techniques of interest to electrical engineers. N. L. JOHNSON

Design for Decision. By IRWIN D. J. BROSS. New York: The Macmillan Company. 1953. Pp. viii + 276. 30*s*.

This book provides enjoyable reading and is an honest attempt to present current statistical ideas to a wide public. The author has taken considerable trouble to produce an interesting, popular text without hiding difficulties inherent in his subject. It is to be hoped that administrators, and scientists in general, will read this book with serious attention. Should this come to pass the author will have performed a valuable service to statistics in making its present-day bases, methods and capabilities known in much wider circles than is now the case.

Decision theory, as is well-known in statistical circles, is the name given to the system developed by Wald in which he gave more formal emphasis to prior distributions and loss functions than did the Neyman-Pearson theory. Information on prior distributions and loss functions is often sparse and this had discouraged development of theory and still hampers practical applications. The author recognizes this difficulty and devotes considerable space to discussion of 'Probability Systems' and 'Value Systems', i.e. the establishment of prior distributions and loss functions respectively. This discussion is followed by an interesting chapter on 'Rules of Action'. Further chapters are entitled 'Operating a Decision-Maker', 'Sequential Decision', 'Data', 'Models' and 'Sampling'. The book might be improved if, with some rearrangement of material, it stopped at this point, except for the excellent final chapter on general aspects of decision problems. The remaining chapters tend to spoil the general picture by over-elaboration. 'Measurement' can be bettered by many elementary statistical texts; 'Statistical Inference' is an inadequate account of the classical theory of tests and estimation implying an artificial distinction between Inference-Makers and Decision-Makers; while the penultimate chapter on 'Statistical Techniques' is competent but dull and necessarily very compressed.

Although the book is written for popular consumption, professional statisticians should find it of very real value. It should stimulate renewed, and serious, thought about the foundations of their subject, and about their responsibilities in fashioning an instrument which may possibly affect deeply the mental outlook of their fellows in years to come. It is also an excellent antidote to those overdoses of mathematical trivia which we can at present hardly avoid.

N. L. JOHNSON

Statistical Theory in Research. By R. L. ANDERSON and T. A. BANCROFT. New York: McGraw-Hill Book Co. Ltd. 1952. Pp. xviii + 389. 59*s*. 6*d*.

Among the many statistical text-books now pouring from the printing presses of many countries this book written by R. L. Anderson and T. A. Bancroft is outstanding. The authors divide it into two parts; the first which they entitle 'Basic Statistical Theory' is just what it says. It is a competent exposition of the corpus of statistical theory which we might now term orthodox and consists of material which has been expounded and mis-expounded by many authors.

The second part entitled 'Analysis of Experimental Models by Least Squares' is interesting, useful, informative and stimulating. It contains chapters on regression analysis, regression with variates, curvilinear regression, complete and incomplete blocks, factorial experiments, analysis of variance components, and discussion of the analysis of data when the model is a mixture of random and systematic effects. Various computational methods including the inversion of matrices are included.

The treatment throughout is unified and mathematical clarity is achieved. What is somewhat surprisingly linked with theoretical clearness is adequate arithmetical exposition of problems allied to theory. The book is to be recommended to all students following a course of analysis of variance and experimental design and to those working statisticians who carry out arithmetical operations every day and who would like to know more of the theoretical background of their operations.

F. N. DAVID

Advanced Statistical Methods in Biometric Research. By C. R. RAO. New York: John Wiley and Sons Inc.; London: Chapman & Hall. 1952. Pp. xvii + 383. 60*s*.

This book, one of a useful series published by John Wiley, is refreshing to read after the number of books with more or less stereotyped outlook which have been published during the years since the war. Dr Rao has confined himself for illustrative purposes to the field of biometry and in particular to the field of anthropometry, but he has actually written a text-book of statistics which is interesting and to a certain extent novel in exposition.

The first thirty pages cover the development of the matrix algebra, an understanding of which is necessary for the development of multivariate analysis. There follow chapters covering distribution theory, the testing of hypotheses, the theory of estimation and the method of maximum likelihood, all of which are familiar, and which have been covered by other authors. It is only when some two-thirds through the book that Dr Rao begins the development of the theory to which the Indian school of statisticians has made such striking contributions. Most research workers nowadays have at least a nodding acquaintance with the problems arising from multivariate analysis, Mahalanobis's D^2 criterion and the discriminant functions of R. A. Fisher. Instead of having to search statistical literature for the exposition of these various statistical techniques, they will now find them gathered together and given coherence in the last three chapters of Dr Rao's book. Probably all workers in the biometric field can read these chapters with profit, even if many will ultimately shirk the arithmetic and turn to Penrose's work.

Having said, or implied, that this is a useful book the reviewer would make one pointed criticism. The theme of the book is an extract by R. A. Fisher: 'The value of the result (to the statistician) follows solely from the value of the material given him. It contains so much information and no more. His job is only to produce what it contains.' On first reading this one thinks how true. But, and this is important, the statistician automatically puts more into the material than it contains every time he makes the basic assumption of normality. Sometimes this assumption does not matter; the reviewer has the suspicion that in multivariate tests it may on occasion matter very much, and certainly insufficient emphasis is laid on this possibly important point.

The book is reasonably well provided with references to original literature with the possible exception of researches carried out by students of Neyman and Pearson. For example, Kolodzieczyk, who did not survive the last war and who made an important contribution to the theory of testing linear hypotheses, merits more than an end-of-chapter reference. Readers will be able to supply other omissions for themselves.

F. N. DAVID

Sampling Methods for Censuses and Surveys, 2nd edition. By F. YATES. London: Charles Griffin and Co. Ltd. 1953. Pp. xvi + 401. 38*s*.

The first edition of this book immediately became the standard manual for survey practitioners. Its practical detail and many worked examples have also made it very useful to teachers, although it makes no attempt to develop the theory of the subject, and therefore requires supplementation by a text such as that recently published by Prof. Cochran.

This new edition leaves unchanged the earlier text, but adds two new chapters. Chapter 9 is devoted to methods of critical analysis of surveys of an investigational character, special attention being paid to the treatment of contrasts between domains of study, which may cut across the original stratification of the population. This important problem is commonly encountered, and there will be a general welcome for its inclusion among the new material.

Chapter 10 is a collection of miscellaneous developments, largely devoted to bringing up to date the exposition in the first edition. The fact that the index has been extended to cover the new material means that the fragmentary nature of Chapter 10 will not prevent its being of use even to those not familiar with the main body of the book.

The select bibliography which was a feature of the first edition has been supplemented, and, for the benefit of those new to the subject, a short course of reading in the book has been indicated.

It is a tribute to the quality of the first edition that no competitor has appeared; despite the jump in price from 24*s*. to 38*s*., none is likely to rival this second edition.

ALAN STUART

Stochastic Processes. By J. L. DOOB. New York: John Wiley and Sons Inc.; London: Chapman and Hall. 1953. Pp. vi + 65. 80*s*.

Since the birth of modern probability occurred with the fundamental paper by Kolmogoroff, the development of the mathematics of probability has been almost exclusively the work of the French and Russian schools, or of scholars who have received their basic training in these schools. The principal exception to this French-Russian monopoly is J. L. Doob, and probabilists everywhere will read his book with interest. Prof. Doob takes as his foundation the axiomatic definition of Kolmogoroff and probability to him is a branch of the theory of Lebesgue measure. Starting from here, what Prof. Doob has really written is a text-book covering the mathematical development of probability theory, the

emphasis being on stochastic processes because it is here that such rapid strides have been and are being made. It is in this field, incidentally, that Prof. Doob himself has made such distinguished contributions.

The book will be of little use to anyone who has not been trained in modern methods of mathematical analysis; it will be no use at all to the non-mathematician seeking for applications. For the mathematical statistician seeking to make rigorous his knowledge of probability theory and wishing to learn up to the boundary of present day research it will be indispensable. It is definitely a specialist's book, which is obviously what the author intended, and within the limited field it will be without rival for years to come. Statisticians tend to use loose mathematical arguments—chiefly because it is not possible to solve their problems any other way—but it is well that there are probabilists such as Prof. Doob stressing the formal mathematical arguments in what is, after all, the root and branch of our work.

The price is perhaps an inhibiting factor from the British point of view. Nevertheless new and old workers alike in the probability field will find it a necessity for their personal libraries. F. N. DAVID

Cambridge Elementary Statistical Tables. By D. V. LINDLEY and J. C. P. MILLER. London: Cambridge University Press. 1953. Pp. 36. 5*s*.

The page size of these tables is $11 \times 8\frac{1}{2}$ in. The cover is of strong paper. Priced at 5*s*., they will be considered to be good value by many occasional users of statistical methods and by students taking elementary courses in statistics at school and at the university.

About half of the space is allotted to a table of values of x^2, $\sqrt{x}$, $\sqrt{(10x)}$, x^{-1}, $x^{-\frac{1}{2}}$, $(10x)^{-\frac{1}{2}}$ and $\log x$ for $x = 1{\cdot}00$ by intervals of $0{\cdot}01$ to $x = 10{\cdot}00$. In general, four-figure accuracy is given and first differences are printed. There is also an antilogarithm table and a table of $\log n!$ up to $n = 300$.

Most of the remaining space is devoted to the common statistical distribution functions. Areas under the normal curve are given in general to four figures against values of the argument proceeding by intervals of $0{\cdot}01$. There are also very brief tables of normal ordinates and percentage points. Then there are tables of percentage points of t, χ^2 and F. Other short miscellaneous tables complete the booklet.

Collections like the present one naturally invite comparison with larger collections such as those of Fisher and Yates and of Hald, and with the *Biometrika Tables for Statisticians*, recently revised and edited by Pearson and Hartley. These larger volumes differ either by providing more figures in the basic tables, or by providing many specialized tables designed for specific statistical problems, or by including much explanatory matter to illustrate the statistical applications. There is little point, however, in attempting to criticize the present Cambridge publication on the ground that it does not contain much useful matter that is found in these larger collections. Granted the size and the price the authors have in the opinion of the reviewer made an excellent choice of what should be included. The printing is clear and the booklet easy to handle. It should have a large circulation among students and among those practising scientists whose work requires some acquaintance with statistical methods.

B. L. WELCH

PUBLICATIONS OF U.S. DEPARTMENT OF COMMERCE,
NATIONAL BUREAU OF STANDARDS

(i) Tables of normal probability functions. Applied Mathematics Series 23. 1953. Pp. 344. Price $2.75.

These tables were originally issued in 1942 as no. 14 in the Mathematical Tables Project; the present edition incorporates certain corrections to errors discovered in the arguments which had been already included in the 1948 edition.

The main Table 1 contains values computed to 15 decimal places of the two functions

$$Q(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}, \quad P(x) = \frac{1}{\sqrt{(2\pi)}} \int_{-x}^{x} e^{-\frac{1}{2}\alpha^2} d\alpha$$

for $x = 0{\cdot}0000\,(0{\cdot}0001)\,1{\cdot}0000\,(0{\cdot}001)\,7{\cdot}800$. For larger values of x, the functions converge rapidly to 0 and 1, respectively, and their values to 15 decimals may be read by inspection from the last page of the table.

A supplementary table (Table II), covering four pages, gives values of $Q(x)$ and of $1-P(x)$ to seven *significant figures* in the range

$$x = 6{\cdot}00\,(0{\cdot}01)\,10{\cdot}00.$$

A short Introduction by A. N. Lowan describes methods of direct and inverse interpolation and ends with a Bibliography of some of the more important existing tables of normal probability functions.

(ii) Table of natural logarithms for arguments between zero and five to sixteen decimal places. Applied Mathematics Series 31. 1953. Pp. 601. Price $3.25.

This table is a reissue of vol. 3 of the Mathematical Tables Project, Table 10, first published in 1941. It gives to 16 decimal places values of $\log_e x$ for $x = 0{\cdot}0000\,(0{\cdot}0001)\,5{\cdot}0000$. No revisions in the original tabular content have been found necessary. There is a short Introduction by A. N. Lowan.

From the prefatory remarks it is understood that a further volume will complete the reissue, giving values of $\log_e x$ for x in the range 5–10.

E. S. PEARSON

50–100 Binomial Tables. By Harry G. Romig. New York: Wiley; London: Chapman and Hall. 1953. Pp. xxvii + 172. 32*s*.

These tables were prepared while the author was associated with the Bell Telephone Laboratories and issued in 1947 in a preliminary draft. They give to six decimal places:

(*a*) the individual terms ${}^nC_x p^x (1-p)^{n-x}$;

(*b*) the cumulative sum of terms (probability of x or less) of the binomial expansion $(q+p)^n$, where $q = 1-p$, for the range of arguments

$$n = 50\,(5)\,100, \quad p = 0{\cdot}01\,(0{\cdot}01)\,0{\cdot}50.$$

The tables break new ground in the sense that fresh computation was required, whereas the National Bureau of Standards Tables (Applied Mathematics Series 6, 1950) which covered the range $n = 2\,(1)\,49$, were derivable from Karl Pearson's *Tables of the Incomplete Beta-function*.

The author gives a helpful Introduction indicating, among other properties, the relation between the sums of binomial terms and the Incomplete Beta-function; this relation makes it possible to obtain from the present tables values of the latter function falling well beyond the range of Karl Pearson's tables.

A companion volume is promised, covering in detail various methods of interpolation that may be needed to make best use of the material provided. A summary is given of the most commonly needed interpolation formulae. Since the argument interval for n is 5, exact formulae for obtaining individual or cumulative terms of binomials with argument $n \pm 1$, $n \pm 2$ in terms of those for n are of course needed and, in fact, take a simple form.

E. S. PEARSON

(*All Rights reserved*)

BIOMETRIKA. Vol. 41, Parts 1 and 2

CONTENTS

	PAGE
Distribution theory of two estimates for standard deviation based on second variant differences. By A. R. KAMAT	1
The statistical treatment of mean deviation. By J. H. CADWELL	12
Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. By G. S. JAMES	19
The use of the Hankel transform in statistics. I. General theory and examples. By R. D. LORD	44
A note on the consistency and maxima of the roots of likelihood equations. By K. C. CHANDA	56
Grouping methods in the fitting of polynomials to equally spaced observations. By P. G. GUEST	62
New techniques for the analysis of absenteeism data. By A. G. ARBOUS and H. S. SICHEL	77
On the superposition of renewal processes. By D. R. COX and WALTER L. SMITH	91
Continuous inspection schemes. By E. S. PAGE	100
On Nahordnung and Fernordnung in samples of literary texts. By WILHELM FUCKS	116
A distribution-free k-sample test against ordered alternatives. By A. R. JONCKHEERE	133
Two-stage procedures for estimating the difference between means. By S. G. GHURYE and HERBERT ROBBINS	146
A bivariate generalization of Student's t-distribution, with tables for certain special cases. By CHARLES W. DUNNETT and MILTON SOBEL	153
A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance. By ROBERT E. BECHHOFER, CHARLES W. DUNNETT and MILTON SOBEL	170
Inequalities for the normal integral including a new continued fraction. By L. R. SHENTON	177
A confidence region for the solution of a set of simultaneous equations with an application to experimental design. By G. E. P. BOX and J. S. HUNTER	190
On the moments of order statistics in samples from normal populations. By H. RUBEN	200
Statistical treatment of censored data. Part I. Fundamental formulae. By F. N. DAVID and N. L. JOHNSON	228
Simplified decision functions. By G. A. BARNARD	241
MISCELLANEA	
Sequential tests for binomial and exponential populations. By F. J. ANSCOMBE and E. S. PAGE	252
Tables of generalized k-statistics. By S. H. ABDEL-ATY	253
An angular transformation for the serial correlation coefficient. By G. M. JENKINS	261
A distribution arising in the study of infectious diseases. By J. O. IRWIN	266
A note on contagious distributions. By H. R. THOMPSON	268
A general expression for the mean in a simple stochastic expidemic. By H. W. HASKEY	272
Some remarks on confidence or fiducial limits. By THEODORE E. STERNE	275
A note on the probability integral of the correlation coefficient. By B. I. HARLEY	278
A simplified expression for the variance of the χ^2-function on a contingency table. By REED B. DAWSON, Jr.	280
REVIEWS	
D. MAINLAND's 'Elementary Medical Statistics'	281
F. N. DAVID's 'A Statistical Primer'	281
D. A. BELL's 'Statistical Methods in Electrical Engineering'	282
IRWIN D. J. BROSS's 'Design for Decision'	283
R. L. ANDERSON and T. A. BANCROFT's 'Statistical Theory in Research'	283
C. R. RAO's 'Advanced Statistical Methods in Biometric Research'	283
F. YATES's 'Sampling Methods for Censuses and Surveys'	284
J. L. DOOB's 'Stochastic Processes'	284
D. V. LINDLEY and J. C. MILLER's 'Cambridge Elementary Statistical Tables'	285
NATIONAL BUREAU OF STANDARDS, Publications of the U.S. Department of Commerce, Applied Mathematics Series 23, 31	285
HARRY G. ROMIG's '50–100 Binomial Tables'	286

Originally printed in Great Britain at the University Press, Cambridge (Brooke Crutchley, University Printer)
Reprinted by Photo-Lithography by Warren and Son Ltd., Winchester.

PAGE
1
12
19
44
56
62
77
91
100
116
133
146
153
170
177
190
200
228
241
252
253
261
266
268
272
275
278
280
281
281
282
283
283
283
284
284
285
285
286